Received 19 March 2022
Revised 2 April 2022
Accepted 2 April 2022

How to make recommendation systems fair: an adequate utility-based approach

Roengchai Tansuchat
Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand, and
Olga Kosheleva
Department of Teacher Education, The University of Texas at El Paso, El Paso, Texas, USA

Abstract

Purpose – In user-oriented websites, e.g. in news websites or in seller websites, it is important to take the user's preferences into account when deciding which items to place in higher-exposure locations. The traditional approach to solving this problem, based on maximizing the average user utility, leads to unfair solutions, and this eventually hurts the company's bottom line. Because of this, researchers have proposed complex schemes that explicitly add fairness to the formulation of this problem. But since utilities already describe human preferences, it is strange that it is necessary to add something beyond utilities.

Design/methodology/approach – In this paper, the authors analyze the problem of selecting exposure level for different items from the viewpoint of decision theory, the basic theory underlying all our activities, including economic ones.

Findings – The authors show that a more adequate use of utilities, namely, taking into account that Nash's bargaining solution is a proper way to make group decisions, not maximizing average utility, already leads to fair solutions.

Originality/value – The idea to apply Nash's bargaining solution to the problem of assigning exposure level to different items is new, as well as the analysis that shows that this application restores the fairness, which is missing in the current solutions.

Keywords Recommendation systems, Fairness, Utility

Paper type Research paper

1.Formulation of the problem

1.1 Need to select levels of exposure

A usual user-oriented website – whether it is a news website or a website of a seller – contains a large amount of information. It is not possible to have all this information on the same page, and even on the main page, some items are listed first, some second, etc. In other words, inevitably different items have different levels of exposure from very high to very low.

It is therefore necessary to select which items should receive high exposure and which not. To make this decision, we can use information about what different users prefer. The problem is that different users have different preferences. So, it is necessary to take all these preferences into account when selecting levels of exposure for different items.

Asian Journal of Economics and Banking
Vol. 6 No. 3, 2022
pp. 308-313
Emerald Publishing Limited
e-ISSN: 2633-7991
p-ISSN: 2615-9821
DOI 10.1108/AJEB-03-2022-0031

1.2 Current utility-based approach to selecting levels of exposure

Since we are talking about user-oriented systems, a natural idea is to follow the users' preferences, and in decision theory, such preferences are described by numerical values known as *utilities*; see, e.g. Fishburn (1969), Fishburn (1988), Luce and Raiffa (1989), Raiffa (1997), Nguyen *et al.* (2009), Nguyen *et al.* (2012) and Kreinovich (2014). Each user i makes a decision that maximizes his/her utility u_i.

To combine these utilities, a seemingly natural idea is

(1) To add the utilities of all the users and

(2) To use maximizing this sum as a criterion for selecting levels of exposure.

Comment. This is similar to how we gauge the state of a country's economy – crudely speaking, we add up all the incomes and consider the resulting sum – gross domestic product (GDP) – as an appropriate measure for comparing different countries.

1.3 Limitations of the current approach

To illustrate what is wrong with this seemingly reasonable idea, let us consider a simplified example of a news website that serves both left- and right-leaning users; this example is, in effect, borrowed from Singh and Joachims (2018) and Joachims *et al.* (2021). Every day, there are some left-leaning new articles and some right-leaning new articles, and it is known that each user prefers the news articles that are closer to his/her own beliefs.

On top of the news website, we can place

(1) Either a link to a left-learning new article

(2) Or a link to a right-leaning news article.

What is the best proportion of times p_L when the left-leaning article is placed on top?

To answer this question, let us introduce some notations. Let us denote

(1) By U, the average utility of the user seeing a link to his/her preferred article placed on top and

(2) By u ($u < U$), the average utility of having to go through other links first in order to access the desired link.

Let us also denote

(1) The number of left-leaning users by n_L and

(2) The number of right-leaning users by n_R.

Now, we are ready to perform the analysis:

(1) A left-leaning user:

- Gains utility U in p_L cases and
- Gains utility u in the remaining $1 - p_L$ cases.

Thus, this user's expected utility is equal to $p_L \cdot U + (1 - p_L) \cdot u$.

(1) Similarly, a right-leaning user:

- Gains utility u in p_L cases and
- Utility U in the remaining $1 - p_L$ cases.

Thus, this user's expected utility is equal to $p_L \cdot u + (1 - p_L) \cdot U$.

So, the sum of the utilities of all the users is equal to

$$n_L \cdot (p_L \cdot U + (1 - p_L) \cdot u) + n_R \cdot (p_L \cdot u + (1 - p_L) \cdot U). \tag{1}$$

One can see that this expression is a linear function of the unknown p_L. We can explicitly describe this expression as a linear function:

$$p_L \cdot (n_L \cdot (U - u) - n_R \cdot (U - u)) + (n_L \cdot u + n_R \cdot U) =$$
$$p_L \cdot (n_L - n_R) \cdot (U - u) + (n_L \cdot u + n_R \cdot U). \tag{2}$$

We want to find the value $p_L \in [0, 1]$ that maximizes the objective function (2). A linear function attains its largest value on an interval at one of its endpoints, i.e. in this case, either for $p_L = 0$ or for $p_L = 1$. In our case, we can see that

(1) When $n_L > n_R$, the maximum is attained when $p_L = 1$, i.e. when the left-leaning article is always placed on top, and

(2) When $n_L < n_R$, the maximum is attained when $p_L = 0$, i.e. when the right-leaning article is always placed on top.

In both cases, this does not seem fair to the minority group that its articles are never placed on top.

Not only it is not fair but it also harmful to the business: when the minority group feels discriminated by this website, they will stop using it and form their own news website, which is, by the way, what often happens.

1.4 What is currently proposed to overcome this limitation

A usual opinion in the recommender community is that the above limitation shows that we must go beyond utility and explicitly take fairness into account when selecting levels of exposure; see, e.g. Joachims *et al.* (2021).

1.5 What we show in this paper

In this paper, we show that the problem is caused not so much by restricting ourselves to utility but rather by an inadequate way utilities of different users are combined now.

We show that if we use an adequate way to combine utilities, then a purely utility-based approach already leads to a fair selection. So there is no need to additionally take fairness into account.

2. How to adequately use utilities

2.1 Utility-based decision-making: reminder

According to the decision theory, a proper way to select an alternative is to maximize the product of utilities $u_1 \cdot \ldots \cdot u_n$; this was proven by the Nobelist John Nash and is thus called *Nash's bargaining solution*; see, e.g. Nash (1950) and Luce and Raiffa (1989).

2.2 How this applies to the above example

In the above two-group example, according to Nash's criterion, we must select the proportion p_L that maximizes the following product:

$$(p_L \cdot U + (1 - p_L) \cdot u)^{n_L} \cdot (p_L \cdot u + (1 - p_L) \cdot U)^{n_R}. \tag{3}$$

Since logarithm is a strictly increasing function, maximizing the product (3) is equivalent to maximizing its logarithm:

$$n_L \cdot \ln(p_L \cdot U + (1 - p_L) \cdot u) + n_R \cdot \ln(p_L \cdot u + (1 - p_L) \cdot U). \quad (4)$$

Differentiating the expression (4) with respect to the unknown p_L and equating the derivative to 0, we conclude that

$$\frac{n_L \cdot (U - u)}{p_L \cdot U + (1 - p_L) \cdot u} - \frac{n_R \cdot (U - u)}{p_L \cdot u + (1 - p_L) \cdot U} = 0,$$

i.e. equivalently,

$$\frac{n_L \cdot (U - u)}{p_L \cdot U + (1 - p_L) \cdot u} = \frac{n_R \cdot (U - u)}{p_L \cdot u + (1 - p_L) \cdot U}.$$

Dividing both sides by $U - u$ and inverting the resulting fractions, we get

$$\frac{p_L \cdot U + (1 - p_L) \cdot u}{n_L} = \frac{p_L \cdot u + (1 - p_L) \cdot U}{n_R}.$$

Multiplying both sides by $n_L \cdot n_R$, we get

$$n_R \cdot (p_L \cdot U + (1 - p_L) \cdot u) = n_L \cdot (p_L \cdot u + (1 - p_L) \cdot U).$$

Moving all the terms containing p_L to left side and other terms to the right side, we get

$$p_L \cdot (U - u) \cdot (n_R + n_L) = n_L \cdot U - n_R \cdot u.$$

If we divide both sides by the total population $n_L + n_R$ and take into account that the ratios

$$r_L \stackrel{\text{def}}{=} \frac{n_L}{n_L + n_R} \text{ and } r_R \stackrel{\text{def}}{=} \frac{n_R}{n_L + n_R} = 1 - r_L$$

describe the proportion of the two types of users, then we get

$$p_L \cdot (U - u) = r_L \cdot U - r_R \cdot u,$$

so

$$p_L = \frac{r_L \cdot U - r_R \cdot u}{U - u}. \quad (5)$$

Of course, this formula only makes sense if the right-hand side of formula (5) is between 0 and 1:

(1) If the right-hand side of formula (5) is smaller than 0, then we should take $p_L = 0$, and

(2) If the right-hand side of formula (5) is larger than 1, then we should take $p_L = 1$.

Here

(1) The inequality $0 \leq p_L$ is equivalent to

$$0 \leq r_L \cdot U - r_R \cdot u = r_L \cdot U - (1 - r_L) \cdot u = r_L \cdot (U + u) - u.$$

Thus, this inequality is equivalent to

$$r_L \geq \frac{u}{u + U}.$$

(2) Similarly, the inequality $p_L \leq 1$, i.e.

$$\frac{r_L \cdot U - r_R \cdot u}{U - u} \leq 1,$$

is equivalent to $r_L \cdot U - r_R \cdot u \leq U - u$, i.e. to

$$r_L \cdot U - (1 - r_L) \cdot u = r_L \cdot (U + u) - u \leq U - u.$$

By adding u to both sides of the last inequality, we get $r_L \cdot (U + u) \leq U$, i.e.

$$r_L \leq \frac{U}{u + U}.$$

Thus, we arrive at the following conclusion.

2.3 Conclusion

(1) If the proportion r_L of the L-group is small, namely, smaller than

$$\frac{u}{u + U},$$

then links to messages favored by this group should never appear on top.

(2) If the proportion r_L of the L-group is sufficiently large, namely, larger than

$$\frac{U}{u + U},$$

then links to messages favored by this group should always appear on top.

(3) In all other cases, links to messages favored by this group should be placed on top with frequency

$$\frac{r_L \cdot U - r_R \cdot u}{U - u}.$$

2.4 Examples

(1) In a realistic case when $U = 2u$

- Messages favored by a group smaller than 1/3 will never be on top,
- Messages favored by a group larger than 2/3 should always appear on top and
- In intermediate cases $1/3 \leq r_L \leq 2/3$, messages of both groups should appear on top.

(2) When the groups are almost equal, i.e. when $r_L \approx r_R \approx 0.5$, we have

$$p_L \approx \frac{0.5U - 0.5u}{U - u} = 0.5.$$

In other words,

- In approximately half of the cases, the left-leaning article is on top and
- In approximately half of the cases, the right-leaning article is on top.

This is exactly the fairness that was missing in the current solution.

References

Fishburn, P.C. (1969), *Utility Theory for Decision Making*, John Wiley & Sons, New York.

Fishburn, P.C. (1988), *Nonlinear Preference and Utility Theory*, The John Hopkins Press, Baltimore, MD.

Joachims, T., London, B., Su, Y., Swaminathan, A. and Wang, L. (2021), "Recommendations as treatments", *AI Magazine*, Vol. 42 No. 3, pp. 19-30.

Kreinovich, V. (2014), "Decision making under interval uncertainty (and beyond)", in Guo, P. and Pedrycz, W. (Eds), *Human-Centric Decision-Making Models for Social Sciences*, Springer-Verlag, Berlin, Heidelberg, pp. 163-193.

Luce, R.D. and Raiffa, R. (1989), *Games and Decisions: Introduction and Critical Survey*, Dover, New York, NY.

Nash, J. (1950), "The bargaining problem", *Econometrica*, Vol. 18 No. 2, pp. 155-162.

Nguyen, H.T., Kosheleva, O. and Kreinovich, V. (2009), "Decision making beyond Arrow's 'impossibility theorem', with the analysis of effects of collusion and mutual attraction", *International Journal of Intelligent Systems*, Vol. 24 No. 1, pp. 27-47.

Nguyen, H.T., Kreinovich, V., Wu, B. and Xiang, G. (2012), *Computing Statistics under Interval and Fuzzy Uncertainty*, Springer-Verlag, Berlin, Heidelberg.

Raiffa, H. (1997), *Decision Analysis*, McGraw-Hill, Columbus, OH.

Singh, A. and Joachims, T. (2018), "Fairness of exposure in rankings", *Proceedings of the 2018 ACM International Conference on Data Discovery and Data Mining SIGKDD*, London, August 19-23, 2018.

Corresponding author
Olga Kosheleva can be contacted at: olgak@utep.edu