Measuring Individual User Fairness with User Similarity and Effectiveness Disparity

Theresia Veronika Rampisela, Maria Maistro, Tuukka Ruotsalo, Christina Lioma

University of Copenhagen

ECIR 2026, Delft
30 March 2026

ADD algorithms data & democracy

IR LAB
INFERENCE & RETRIEVAL LAB

Individual fairness for users

=

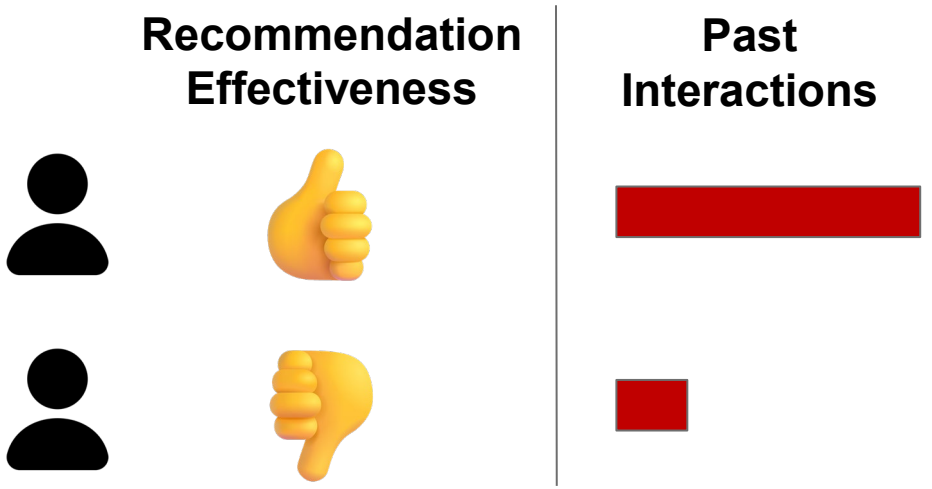
similar treatment for (similar) individual users

Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In ITCS 2012 - Innovations in Theoretical Computer Science Conference. 214–226. https://doi.org/10.1145/2090236.2090255

User-side Individual Fairness in Recommender System: Definition #1

Different Effectiveness = Unfair

User-side Individual Fairness in Recommender System: Definition #1



Different amount of data →
Different performance

Different Effectiveness = Unfair

Problem #1: no consideration for **user similarity** (based on past interactions)

User-side Individual Fairness in Recommender System: Definition #2

Similar Recommendations for **Similar Users = Fair**

User-side Individual Fairness in Recommender System: Definition #2

similar recommendations
≠ equal satisfaction

Similar Recommendations for **Similar Users = Fair**

Problem #2: no consideration for **effectiveness disparity**

New Evaluation Measure for User-side Individual Fairness in Recommender Systems

Fairness Measure	Effectiveness Disparity	User Similarity
Standard deviation	✓	-
Gini index	✓	-
Envy-based measures	✓	-
UF	-	✓
PUF (our measure)	✓	✓

Similar Effectiveness for **Similar Users = Fair**

Our contribution: Pairwise User unFairness (PUF)

$$\text{PUF} = \frac{2}{m(m-1)} \sum_{u \in U} \sum_{u' \in U \setminus \{u\}} sim(u, u') \times |S(u) - S(u')|$$

average across all user pairs

User Similarity

Effectiveness Disparity

Lower PUF score is fairer

HIGH	×	**LOW**	➡	**LOW (FAIR)**
LOW	×	**HIGH**	➡	**LOW (FAIR)**
LOW	×	**LOW**	➡	**LOW (FAIR)**
HIGH	×	**HIGH**	➡	**HIGH (UNFAIR)**

Key Findings from Experimental Results

1. Existing fairness measures **cannot be used as a proxy** for PUF

2. PUF **captures** varying levels of effectiveness and user similarities

3. PUF **correctly** scores close to the fairest for the most fair case

Paper

Thank you!