NDBI021, Lecture 6

User preferences, 2/1 ZK+Z,

Wed 12:20 – 13:50 S8

Wed 14:00 – 15:30 SW2 (odd weeks)

https://www.ksi.mff.cuni.cz/~peska/vyuka/ndbi021/2022/

matfyz

siret research group

https://ksi.mff.cuni.cz

Fairness in Recommender Systems

Tutorial on Fairness of Machine Learning in Recommender Systems

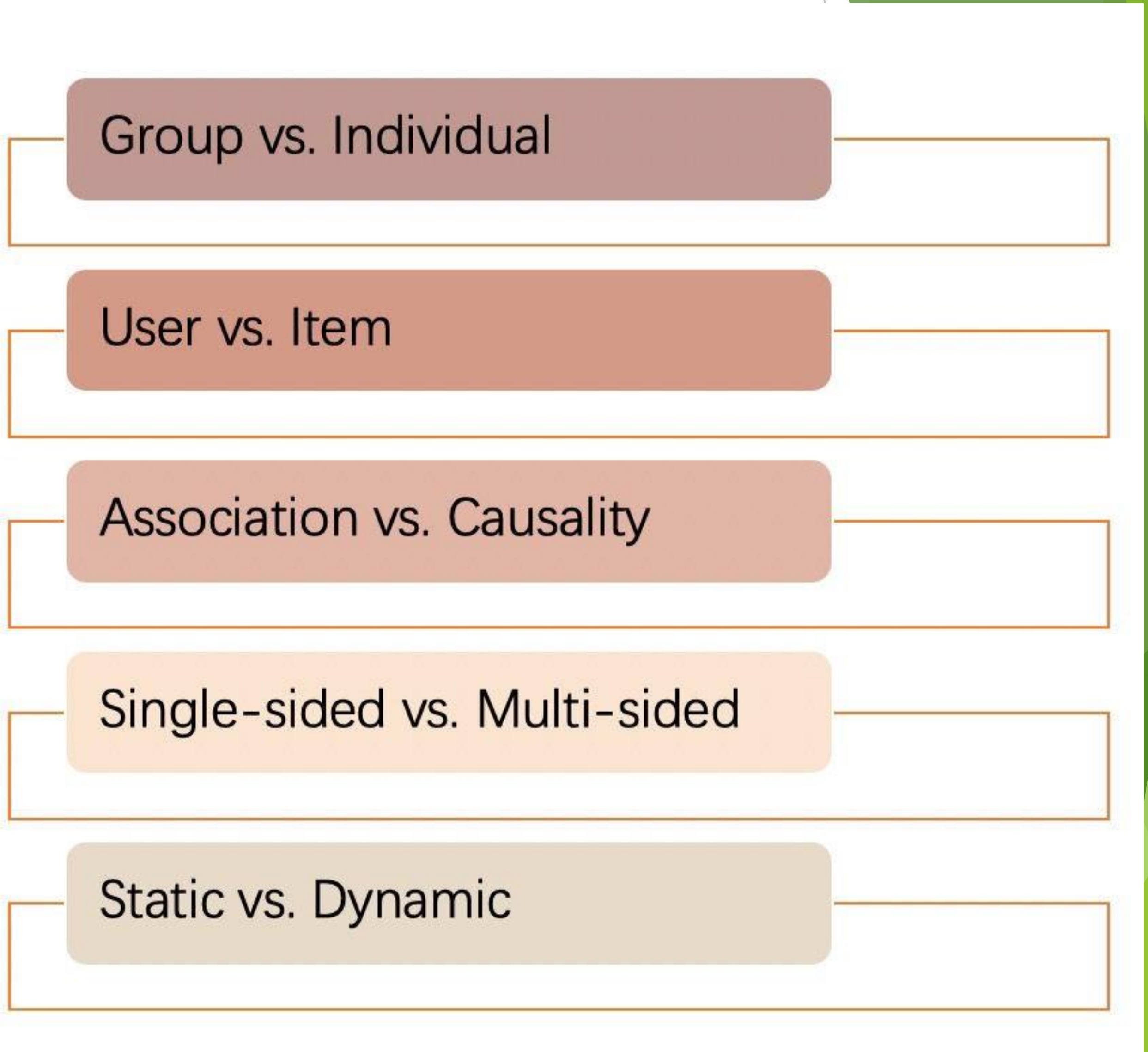
Yunqi Li
Rutgers University
yunqi.li@rutgers.edu

Yingqiang Ge
Rutgers University
yingqiang.ge@rutgers.edu

Yongfeng Zhang
Rutgers University
yongfeng.zhang@rutgers.edu

Tutorial from: https://fairness-tutorial.github.io/

Taxonomies



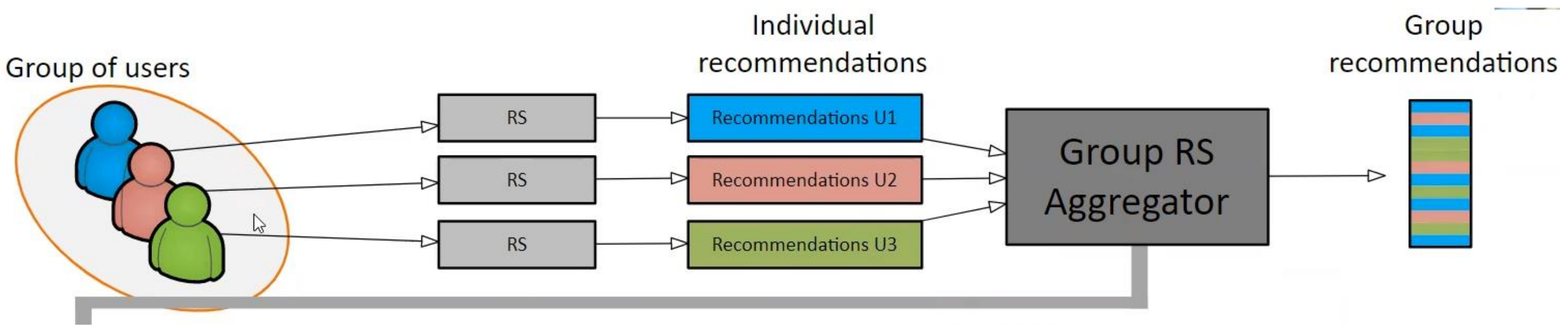
Group vs. Individual
User vs. Item
Association vs. Causality
Single-sided vs. Multi-sided
Static vs. Dynamic

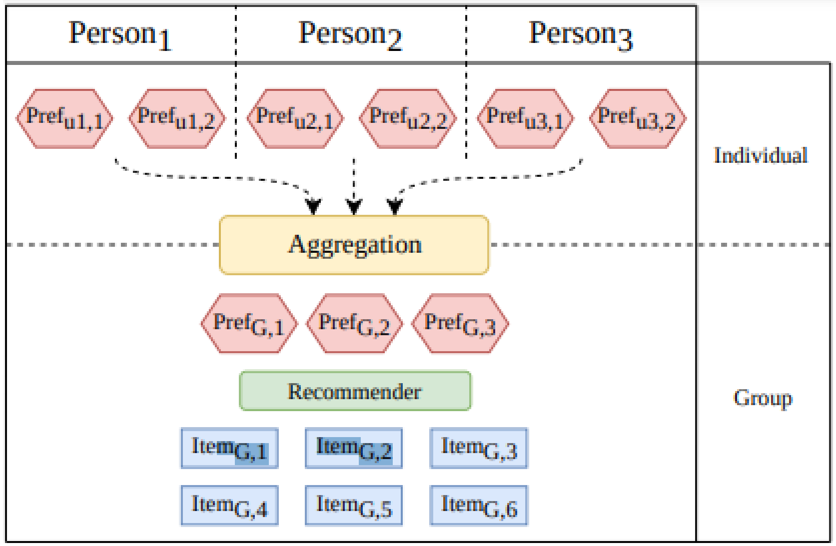
Fairness in RS, further reading

- https://link.springer.com/article/10.1007/s11257-020-09285-1
- https://dl.acm.org/doi/pdf/10.1145/3383313.3411545
- https://www.sciencedirect.com/science/article/pii/S0306457321001503
- https://arxiv.org/abs/1908.06708
- https://dl.acm.org/doi/pdf/10.1145/3450614.3461685
- https://arxiv.org/abs/2006.05255
- https://dl.acm.org/doi/pdf/10.1145/3184558.3186949

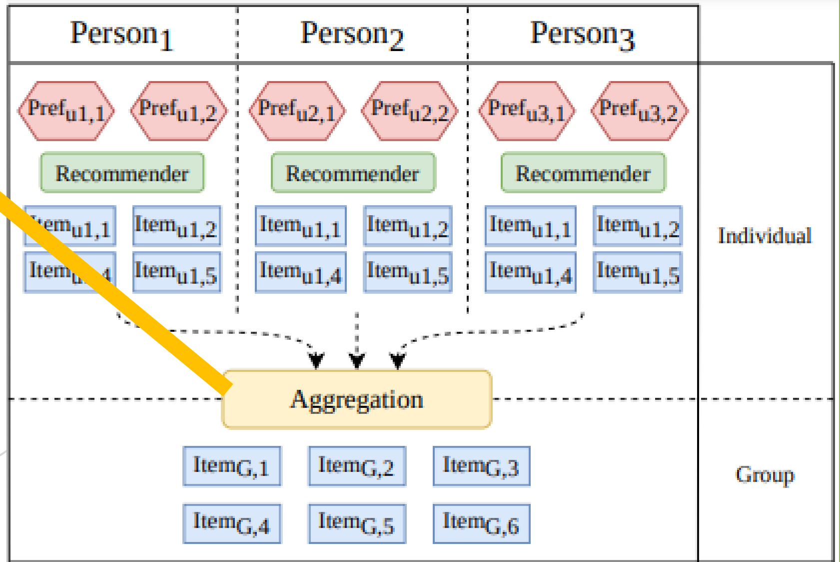
Group Recommender Systems

Group RS





- Two main classes of approaches:
  - Item-wise (utility of each item is evaluated independently from all others)
  - List-wise (utility is evaluated for the whole list)
    - In most cases, the list is constructed incrementally
    - You can focus on ranking-aware fairness, i.e., each head of the list of recommendations should be as fair as possible



Fairness in Group Recommendation

- **Group Recommendation**: recommend items to groups of users whose preferences can be different from each other.
- **Fairness Concerns**: maximize the satisfaction of each group member while minimizing the unfairness (the imbalance of user utilities inside the group) between them.
- **Why not just aggregate individual preferences of users?**
  - **Possible unfairness for individuals with minority opinion**

Xiao, Lin, et al. "Fairness-aware group recommendation with pareto-efficiency." Recsys'17

Item-based Group RS

Many strate

[Masthoff, 2004]

- Average
- Least misery
- Average without misery
- Multiplicative
- Plurality Voting
- Borda count
- Copeland rule
- Approval voting
- Most pleasure
- Fairness
- Most respected person

https://pro.unibz.it/projects/schoolrecsys17/JudithMasthoff.pdf

- **Additive utilitarian** (ADD, consensus)
  Sum of scores for an item across the group

$$\operatorname*{argmax}_{i \in I} \sum_{u \in G} \text{score}(u, i) \tag{4.1}$$

- **Approval Voting** (APP, majority)
  Number users that like the item above a certain threshold

$$\operatorname*{argmax}_{i \in I} \left| \{u \in G : \text{score}(u, i) \geq \text{treshold}\} \right| \tag{4.2}$$

- **Average** (AVG, consensus)
  Average of scores for an item across the group

$$\operatorname*{argmax}_{i \in I} \frac{\sum_{u \in G} \text{score}(u, i)}{|G|} \tag{4.3}$$

- **Average without Misery** (AVM, consensus)
  Average of scores for an item across the group only if the item is above a certain threshold for all group members

$$\operatorname*{argmax}_{i \in I : \nexists u \in G | \text{score}(u,i) \leq \text{treshold}} \frac{\sum_{u \in G} \text{score}(u, i)}{|G|} \tag{4.4}$$

- **Multiplicative** (MUL, consensus)
  Multiplies all received ratings together.

$$\operatorname*{argmax}_{i \in I} \left(\prod_{u \in G} \text{score}(u, i) \right) \tag{4.12}$$

- **Plurality Voting** (PLU, majority)
  Each user has a set number of votes that get distributed. The item with the most received votes is selected.

$$\operatorname*{argmax}_{i \in I} \left(\sum_{u \in G} \text{VotesAwarded}(u, i) \right) \tag{4.13}$$

Where *votes awarded* is some function that decides for each user how the available votes will be distributed among the items.

- **Borda count** (BRC, majority)
  Sum of scores derived from item rankings. The ranking score is defined for each user by ordering the user's items by score and awarding points corresponding to the location of the item in this ordered list. Worst item receiving 1 point and best item $|I|$ points.

$$\operatorname*{argmax}_{i \in I} \left(\sum_{u \in G} \text{RankingScore}(u, i) \right) \tag{4.5}$$

Where *ranking score* is defined as follows:

$$\text{RankingScore}(u, i) := \left| \{ i_{other} \in I : \text{score}(u, i_{other}) \leq \text{score}(u, i) \} \right|$$

- **Copeland rule** (COP, majority)
  Difference between number of wins and losses for pair-wise comparison of all items

$$\operatorname*{argmax}_{i \in I} \left(\text{W}(t, I - t) - \text{L}(t, I - t) \right) \tag{4.6}$$

- **Fairness** (FAI, consensus)
  Users, in turn, one after another select their top item.

$$\operatorname*{argmax}_{i \in I} \text{score}(u_{current}, i) \tag{4.7}$$

Where $u_{current}$ is user selected from G for each iteration according to some (in most cases circular, or ping pong) rule.

- **Least misery** (LMS, borderline)
  Uses the lowest received rating among the group members as the item's aggregated rating.

$$\operatorname*{argmax}_{i \in I} \left(\min_{u \in G} (\text{score}(u, i)) \right) \tag{4.8}$$

- **Most Pleasure** (MPL, borderline)
  Uses the highest received rating among the group members as the item rating.

$$\operatorname*{argmax}_{i \in I} \left(\max_{u \in G} (\text{score}(u, i)) \right) \tag{4.9}$$

- **Majority Voting** (MAJ, majority)
  Uses the rating that was given by the majority of the group's members. (Can only work on discrete ratings)

$$\operatorname*{argmax}_{i \in I} \left(\operatorname*{mode}_{u \in G} (\text{score}(u, i)) \right) \tag{4.10}$$

- **Most Respected Person** (MRP, borderline)
  Uses rating proposed by the most respected member of the group.

$$\operatorname*{argmax}_{i \in I} \text{score}(u_{most\_respected}, i) \tag{4.11}$$

Fairness in Group Recommendation

- **Item-based group RS**
  - **Faster**
  - **Still high chance of unfairness**

Xiao, Lin, et al. "Fairness-aware group recommendation with pareto-efficiency." Recsys'17

Fairness in Group Recommendation

- **Method**:
  - The **Social Welfare** ($SW(g, I)$): overall utility of all users inside the group g given a group recommendation I.
  - The **Fairness** ($F(g, I)$): a function of $U(u, I), \forall u \in g, \forall I$.
  - Multi-Objective Optimization: $\lambda \cdot SW(g, I) + (1 - \lambda) \cdot F(g, I)$
- **Experiment Results**: The results indicate that considering fairness can improve the quality of group recommendation.

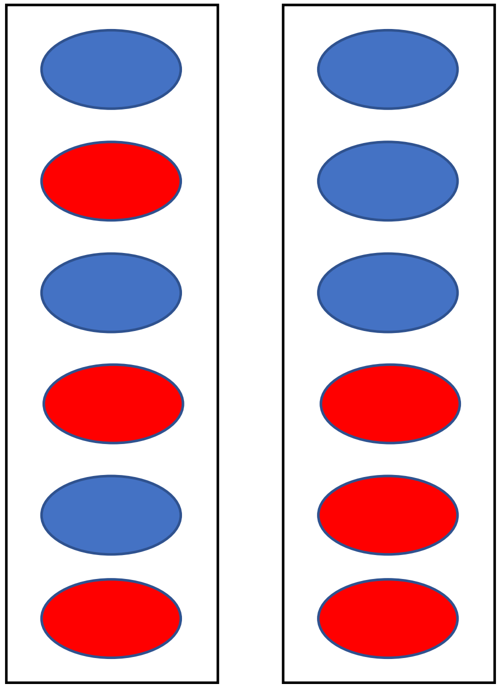
λ, RG	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
F@K	0.0260	0.0817	0.0877	0.0953	0.1019	0.1041	0.1046	0.1053	0.1058	0.1062	0.1062
NDCG@K	0.0697	0.2200	0.2287	0.2334	0.2394	0.2423	0.2440	0.2421	0.2459	0.2478	0.2476

Xiao, Lin, et al. "Fairness-aware group recommendation with pareto-efficiency." Recsys'17

Fairness in Group Recommendation

Possible issues:

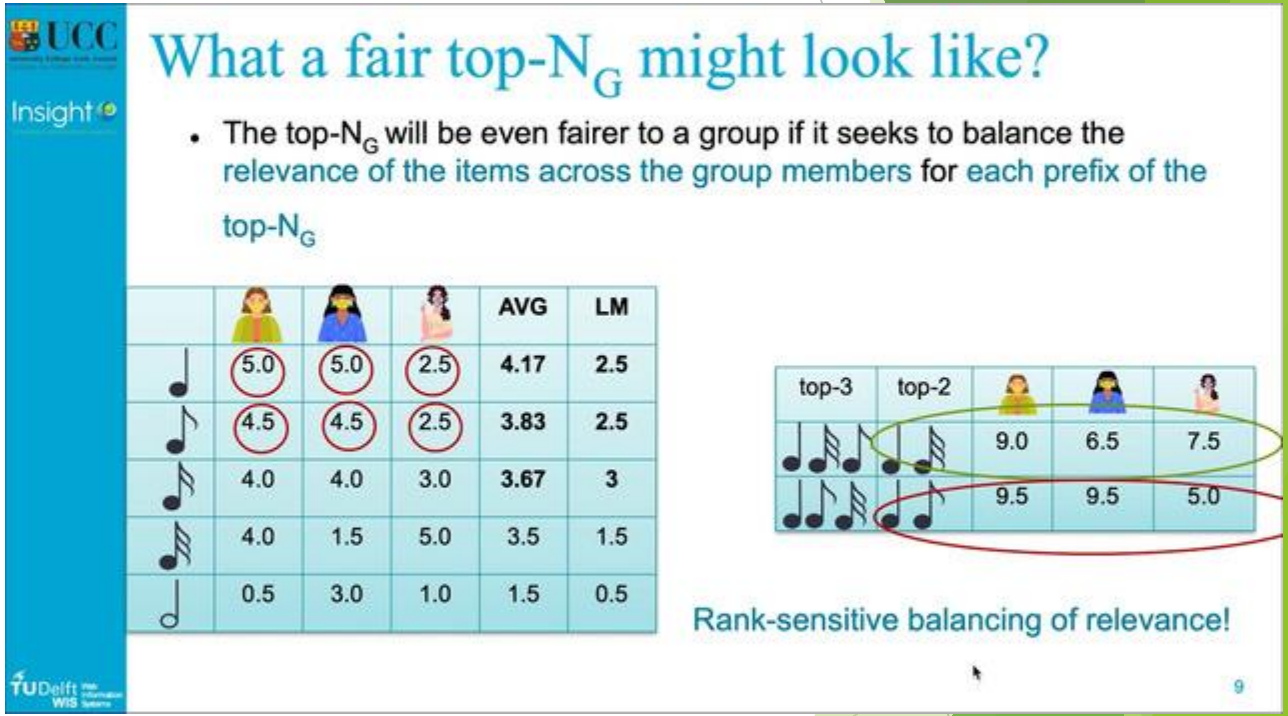
- Fairness metrics does not consider ranking
- User's attention is unevenly distributed



Fairness in Group Recommendation

Possible issues:

- Ranking aware fairness
  - Nice feature: optimizing just one metric comprising both utility and fairness
- Greedy construction algorithm GFAR



https://slideslive.com/38934807/ensuring-fairness-in-group-recommendations-by-ranksensitive-balancing-of-relevance?ref=speaker-41949
https://dl.acm.org/doi/10.1145/3383313.3412232

Fairness in Group Recommendation

Drastic decrease of relevance w.r.t. order of items

Finding one item per user is sufficient

3.1 GFAR's definition of fairness

For a group member $u \in G$, let $p(\text{rel} \mid u, i)$ be the probability that item i is relevant to u. We estimate $p(\text{rel} \mid u, i)$ as:

$$p(\text{rel} \mid u, i) = \frac{\text{Borda-rel}(u, i)}{\sum_{j \in \text{top-}N_u} \text{Borda-rel}(u, j)} \tag{1}$$

Following Xiao et al. [18], we define $\text{Borda-rel}(u, i) = |\{j : \text{rank}(j, \text{top-}N_u) > \text{rank}(i, \text{top-}N_u), \forall_j \in \text{top-}N_u\}|$, where, from above, $\text{rank}(i, \text{top-}N_u)$ is the rank of item i in u's top-N candidate items, which are obtained using the $s(u, i)$ scores predicted by the underlying recommender algorithm.[2]

Let also $p(\neg \text{rel} \mid u, S)$ be the probability that none of the items in set S are relevant to user u. Then, we derive the probability that at least one item within S is relevant to u, $p(\text{rel} \mid u, S)$, as follows:

$$\begin{aligned} p(\text{rel} \mid u, S) &= 1 - p(\neg \text{rel} \mid u, S) \\ &= 1 - \prod_{i \in S} (1 - p(\text{rel} \mid u, i)) \end{aligned} \tag{2}$$

Now, from $p(\text{rel} \mid u, S)$ for each group member $u \in G$, we define $f(S)$ as the sum of each group member's probability of finding at least one relevant item within the set S:

$$f(S) = \sum_{u \in G} p(\text{rel} \mid u, S) = \sum_{u \in G} \left(1 - \prod_{i \in S} (1 - p(\text{rel} \mid u, i))\right) \tag{3}$$

[2] A more obvious definition is $p(\text{rel} \mid u, i) = s(u, i) / \sum_{j \in C} s(u, j)$, where $C \subseteq I$ are the candidate items. Compared to Eq. 1, this did not work well in our experiments. The probable explanation is that it relies too heavily on the actual $s(u, i)$ values, whereas Eq. 1 uses their ordering.

Eq. 3 shows how to 'balance' relevance across the group members for a set. It is not yet rank-sensitive. To make it rank-sensitive, we define the marginal gain in function f that arises when we add a new item to the set S, $f(i, S)$, as:

$$f(i, S) = f(S \cup \{i\}) - f(S) \tag{4}$$

Using Eq. 3 and Eq. 4, we can obtain the following:

$$f(i, S) = \sum_{u \in G} [p(\text{rel} \mid u, i) \prod_{j \in S} (1 - p(\text{rel} \mid u, j))] \tag{5}$$

Then, we can define an ordered set to be fair if there is balance in each prefix of the set. In other words, the first item in the set should, as far as possible, balance the interests of all group members; the first two items taken together must do the same; also the first three; and so on up to N:

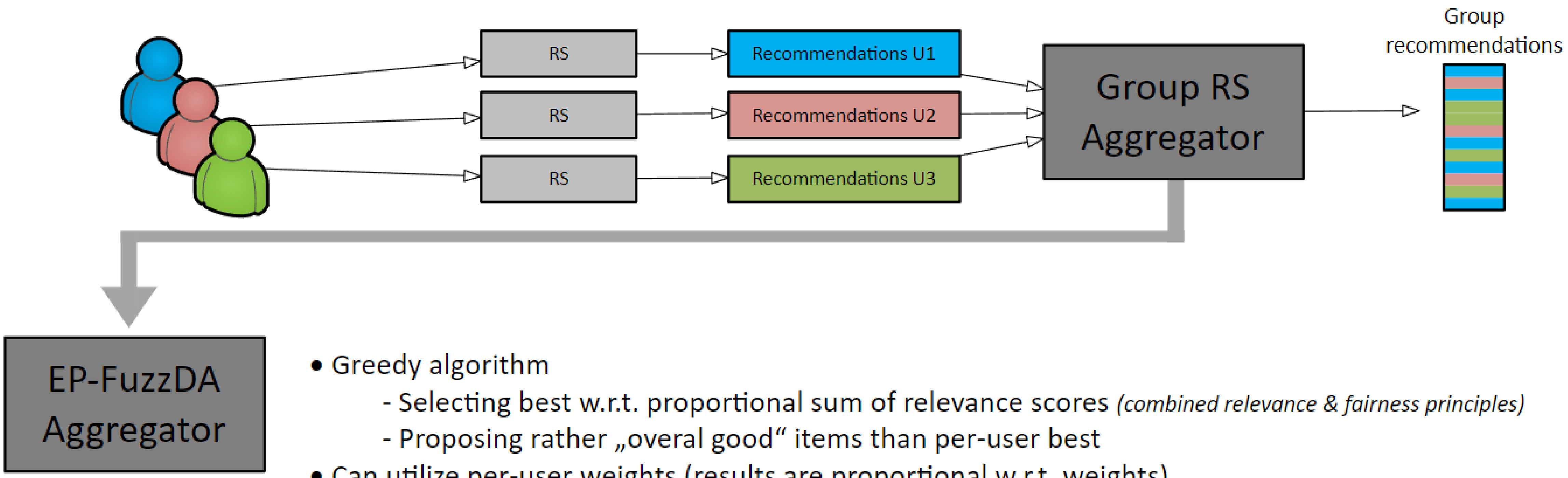
$$\text{fair}(OS) = \sum_{k=1}^{|OS|} f(OS[k], \{i \in OS : \text{rank}(i, OS) < k\}) \tag{6}$$

A natural alternative is to find an approximation of OS^* using a greedy algorithm. The GFAR greedy algorithm starts with an empty set, $OS = \{\}$. At each iteration, it inserts into the ordered result set the item i^* from the remaining candidates (i.e. $C \setminus OS$) that gives the highest marginal gain:

$$i^* = \arg\max_{i \in C \setminus OS} f(i, OS) \tag{8}$$

Fairness-preserving Group Recommendations With User Weighting

Ladislav Malecek, Ladislav Peska



- Greedy algorithm
  - Selecting best w.r.t. proportional sum of relevance scores *(combined relevance & fairness principles)*
  - Proposing rather „overal good“ items than per-user best
- Can utilize per-user weights (results are proportional w.r.t. weights)
  - Suitable for long-term fairness of permanent groups

See the paper for algorithm details & evaluation

Source codes: https://github.com/LadislavMalecek/UMAP2021

Fairness in Group Recommendation

We consider group recommending strategy as fair if all users receive items with approx. the same sum of estimated relevance scores. This statement should hold for all prefixes of the recommendations list.

Input: group members $u \in \mathcal{G}$, andidate items $c \in C$, relevance scores $r_{u,c} \in \mathbf{R}$, number of intems k, user's weights v_u; $\sum v_u = 1$

Output: ordered list of group recommendations L_G^k

$L_G = []$; $TOT = 0$; $\forall u : r_u = 0$
for $i \in [0, ..., k]$ **do**
  for $c \in C \setminus L_G$ **do**
    $TOT_c = TOT + \sum_{\forall u} r_{u,c}$
    $\forall u : e_u = max(0, TOT_c * v_u - r_u)$
    $gain_c = \sum_{\forall u} min(r_{u,c}, e_u)$
  end for
  $c_{best} = \text{argmax}_{\forall c}(gain_c)$; append c_{best} to L_G
  $\forall u : r_u = r_u + r_{u,best}$; $TOT = \sum_{\forall u} r_u$
end for
return L_G

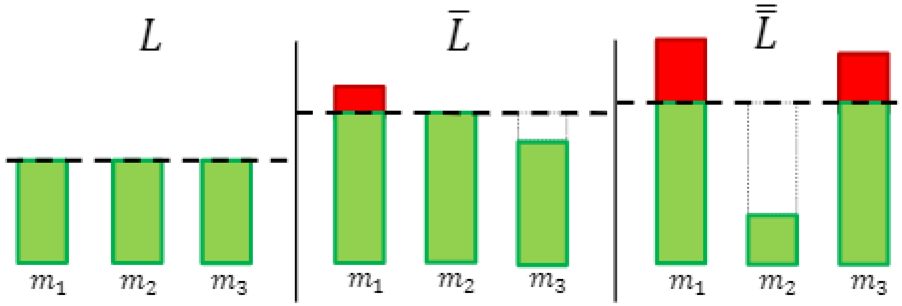


Figure 1: Example of proportionality vs. utility tradeoff. Figure depicts three lists ($L, \bar{L}, \bar{\bar{L}}$) and their score w.r.t. evaluation metrics m_1, m_2, m_3. For simplicity, consider that all metrics has the same weight ($w_1 = w_2 = w_3$). Dashed line denotes exactly proportional fraction of the total utility (i.e. mean utility in the case of equal weights), green bars denote proportional part of metric's utility, while red bar denotes excess over it. L provides perfectly proportional results, but its overall utility is inferior. $\bar{\bar{L}}$ has the highest mean utility (dashed line), but it is highly disproportional. We consider $\bar{L}$ to be the best option as the sum proportional fractions of metric's utility (sum of green bars) is largest.

TOT_c: total relevance of so far recommended objects (plus the relevance of the considered one)

e_u: not yet accounted relevance share of the current user (how much did we ignored this user in the past?)

- **v_u:** weight of individual user. Can e.g. adjust the lack of fairness in previous recommendation sessions

$gain_c$: sum of per-user relevances of considered item (but only the fair portion of per-user relevances are considered)

- for example, if some user is over-represented and his/her $e_u = 0$, relevance w.r.t. this user is completely ignored when calculating the best next object.

Biases in RS

Fairness in evaluation

- Popularity bias (more popular => much more attention)
- Biased historical data (missing not at random) => (unbiased) learning algorithm => biased recommendations
- => biased off-line evaluation (same bias vector => better results)
- => discrepancy between off-line and on-line evaluation

- How to evaluate methods fairly?

Fairness in evaluation

- Inverse propensity score
- Weight results by the inverse to the propensity score
  - (probability of being noticed by the user)
  - Definitions may vary on available information
    - Based on general item's popularity
    - Based on recommended positions
    - Based on user's actions within the page

De-biasing Off-line Evaluation

- https://dl.acm.org/doi/pdf/10.1145/3240323.3240355

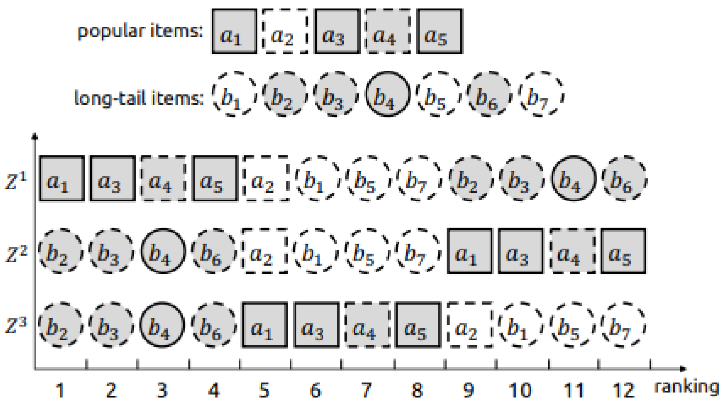


Figure 1: A hypothetical example to illustrate the evaluation bias that results from use of the AOA evaluator. Three recommenders generated distinct lists of recommendations, Z^1, Z^2 and Z^3, for the same user. Among the shaded items that were preferred by the user, the ones with a solid border were observed by recommenders. The performance was measured by DCG, and the results are presented in Table 1.

Table 1: The true and estimated DCG values for three recommenders in Fig. 1. $R(\hat{Z})$ denotes the ground truth, and $\hat{R}_{\text{AOA}}(\hat{Z})$ denotes the AOA estimations. The AOA estimator outputs larger values when popular items are ranked higher.

Estimator	Z^1	Z^2	Z^3
$R(\hat{Z})$	0.463	0.463	0.494
$\hat{R}_{\text{AOA}}(\hat{Z})$	0.585	0.340	0.390

3.1 Average-over-all (AOA) evaluator

In prior literature, $R(\hat{Z})$ was estimated by taking the average over all observed user feedback $\mathcal{S}_u^*$:

$$\hat{R}_{\text{AOA}}(\hat{Z}) = \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}}\frac{1}{|\mathcal{S}_u^*|}\sum_{i\in\mathcal{S}_u^*} c(\hat{Z}_{u,i}) = \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}}\frac{1}{\sum_{i\in\mathcal{S}_u} O_{u,i}}\sum_{i\in\mathcal{S}_u} c(\hat{Z}_{u,i})\cdot O_{u,i} \qquad (6)$$

nDCG, AUC, MAP,...

3.2 Unbiased evaluator

To conduct unbiased evaluation of biased observations, we leverage the IPS framework [16, 22] that weights each observation with the inverse of its propensity, where the term *propensity* refers to the tendency or the likelihood of an event happening. The intuition is to down-weight the commonly observed interactions, while up-weighting the rare ones. In the context of this paper, the probability $P_{u,i}$ is treated as the pointwise propensity score. Therefore, the IPS unbiased evaluator is defined as follows:

$$\hat{R}_{\text{IPS}}(\hat{Z}|P) = \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}}\frac{1}{|\mathcal{S}_u|}\sum_{i\in\mathcal{S}_u^*}\frac{c(\hat{Z}_{u,i})}{P_{u,i}} = \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}}\frac{1}{|\mathcal{S}_u|}\sum_{i\in\mathcal{S}_u}\frac{c(\hat{Z}_{u,i})}{P_{u,i}}\cdot O_{u,i} \qquad (7)$$

Propensity score

$$\hat{P}_{*,i} \propto (n_i^*)^{\gamma}\cdot n_i, \qquad (13$$

where $n_i = \sum_{u\in\mathcal{U}} \mathbf{1}[i \in \mathcal{S}_u]$ and $n_i^* = \sum_{u\in\mathcal{U}, i\in\mathcal{S}_u^*} O_{*,i}$.

However, empirically, n_i is not directly observable. To addres this problem, we observe that n_i^* is sampled from a binomial distri bution[4] parameterized by n_i, that is, $n_i^* \sim \mathcal{B}(n_i, P_{*,i})$. Therefore, a relationship between n_i and n_i^* can be built by bridging the genera tive model (eqn. 13) with the following unbiased estimator:

$$\hat{P}_{*,i} = \frac{n_i^*}{n_i} \propto (n_i^*)^{\gamma}\cdot n_i \qquad (14$$

Therefore, $n_i \propto (n_i^*)^{\frac{1-\gamma}{2}}$. We use this as a replacement for the unobserved n_i in eqn. 13, which results in an unbiased $\hat{P}_{*,i}$ estimato that is determined by only the empirical counts of items:

$$\hat{P}_{*,i} \propto (n_i^*)^{\left(\frac{\gamma+1}{2}\right)} \qquad (15$$

De-biasing Off-line Evaluation

- https://link.springer.com/article/10.1007/s10844-021-00651-y
- Alternative: sampling from test data to de-bias them
  - Based on missing-at-random (MAR) vs. Missing-not-at-random (MNAR)
  - Sample from MNAR data to better resemble MAR
  - **Variants:**
    - You have some subsample that is MAR (random recommendations, forced rating), sample from MNAR so that posterior probability is similar to MAR. Finding weight w for each user-item pair

$$P_{mnar}(u,i|\mathcal{O},w) = P_{mar}(u,i|\mathcal{O}) \quad \forall (u,i) \in D^{mnar}$$

$$P_{mnar}(u|\mathcal{O},w) = P_{mar}(u|\mathcal{O}) \quad \forall u \in U$$

$$P_{mnar}(i|\mathcal{O},w) = P_{mar}(i|\mathcal{O}) \quad \forall i \in I$$

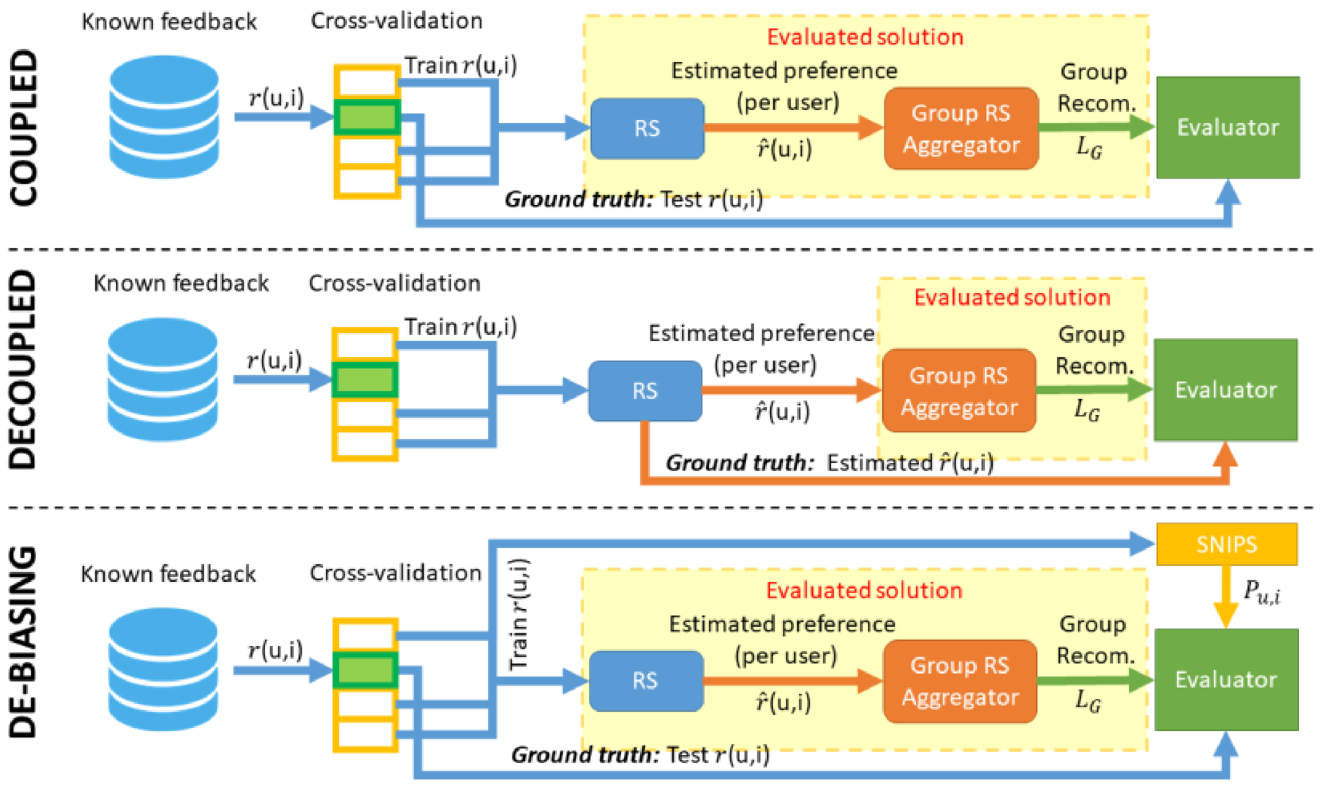
$$w_u = \frac{P_{mar}(u|\mathcal{O})}{P_{mnar}(u|\mathcal{O})} \quad \forall u \in U$$

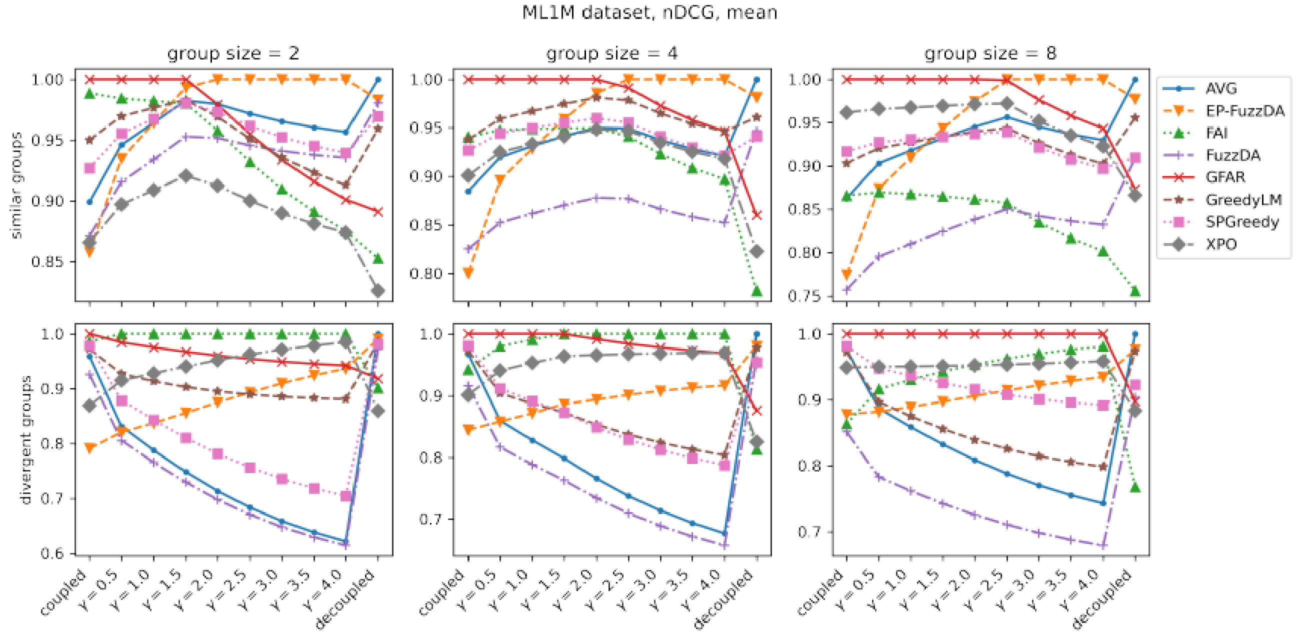
$$w_i = \frac{P_{mar}(i|\mathcal{O})}{P_{mnar}(i|\mathcal{O})} \quad \forall i \in I$$

    - You do not have MAR subsample: assume uniform posterior probability
  - Possible disadvantage: not enough data due to sampling
    - Sample with repetition
  - Possible disadvantage: not enough data from all segments

Biases in metrics

- GFAR vs. FuzzDA – Group RS:
  - What to evaluate for group RS?
  - Decoupled evaluation depends on estimated ratings (their absolute differences)
    - Lower values / higher score differences favor „best-per-user“ algos.
    - Higher values /smaller differences may favor algorithms seeking items best in average
  - Scale [0:10]
  - $\widehat{r_{u,i1}}$ = [4, 7, 3, 4, 6] vs.
    $\widehat{r_{u,i2}}$ = [9, 1, 0, 2, 9]
  - $\widehat{c_{u,i1}}$ = [100, 20, 150, 100, 40] vs.
    $\widehat{c_{u,i2}}$ = [1, 600, 1000, 500, 5]
    - Which one is better?
    - Average estimated relevance vs. Borda count





Biases in metrics

- How to evaluate multiple metrics?
  - Recap: diversity, novelty, popularity bias, relevance

$$div_{sim}(u) = \frac{\sum_{\forall o_i, o_j \in O_u; i \neq j} 1 - sim(o_i, o_j)}{|O_u| * (|O_u| - 1)}$$

$$\text{MMR} = \arg\max_{D_i \in R \setminus S} \left[\lambda \, \text{Sim}_1(D_i, Q) - (1 - \lambda) \max_{D_j \in S} \text{Sim}_2(D_i, D_j) \right]$$

$$IP = \frac{\text{number of users who have rated the item}}{\text{number of users}} \quad (6)$$

An item's novel value (INV) is then measurable by taking the log of the inverse IP:

$$INV = -\log_2(IP) \quad (7)$$

$$DCG_{pos} = rel_1 + \sum_{i=2}^{pos} \frac{rel_i}{\log_2 i}$$

$$PopLift = \frac{mPop_{rec} - mPop_{data}}{mPop_{data}} \quad (13)$$

The $mPop_{rec}$ and $mPop_{data}$ stands for the mean popularity of items that were recommended and items that occurs in the dataset respectively. Formally, suppose to have a list of positive feedback events in a dataset $f_i(u, o) \in \mathcal{F}^+$. Each event is triggered by a user u on an item o. We can use the notation $o_j \in f_i$ meaning that the item o_j is a target in the event f_i. Then popularity of an item is defined as

$$pop(o_j) = \frac{|\{f_i : o_j \in f_i\}|}{|\mathcal{F}^+|}$$

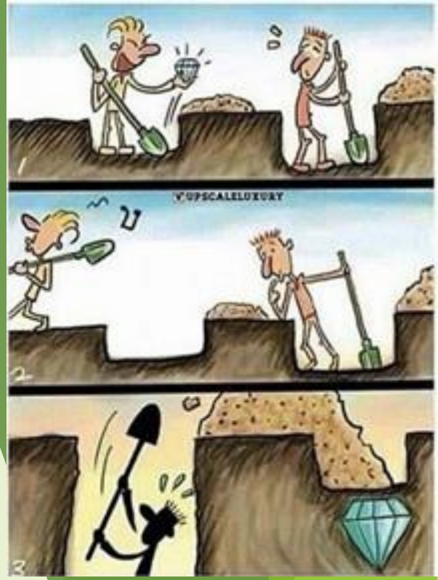
Now, suppose that O_{rec} contains a concatenated list of all recommendations (irrespective of users) and O_{data} contains a list of target items for all events $f_i(u, o) \in \mathcal{F}^+$. Then

$$mPop_{rec} = \frac{\sum_{o_j \in O_{rec}} pop(o_j)}{|O_{rec}|} \text{ and } mPop_{data} = \frac{\sum_{o_j \in O_{data}} pop(o_j)}{|O_{data}|}.$$

Biases in metrics

- How to evaluate multiple metrics?
  - Is it good to trade 0.1 increase in diversity for 0.05 decrease in nDCG?
  - What about methods ranking?
    - But this is affected by the selection of evaluated cases
  - Pareto optimality
    - Hard to find in reality
    - Probabilistic approach: for randomly selected aggregated utility from the set of plausible ones, what is the chance that A1 is better than A2 (idea from https://dsachar.github.io/publication/2019-sac-sac/2019-sac-sac.pdf)
      - Then again, how the plausible set of utilities looks like?

Long vs. short-term evaluation



- Exploration vs. Exploitation tradeoff

  - Purely exploitational RS: high target values in short-term, but possibly low target values in long-term
  - Problematic evaluation
    - No exploration in the train data => no way to learn it => no exploration in the test data => Penalization of exploration-oriented RS

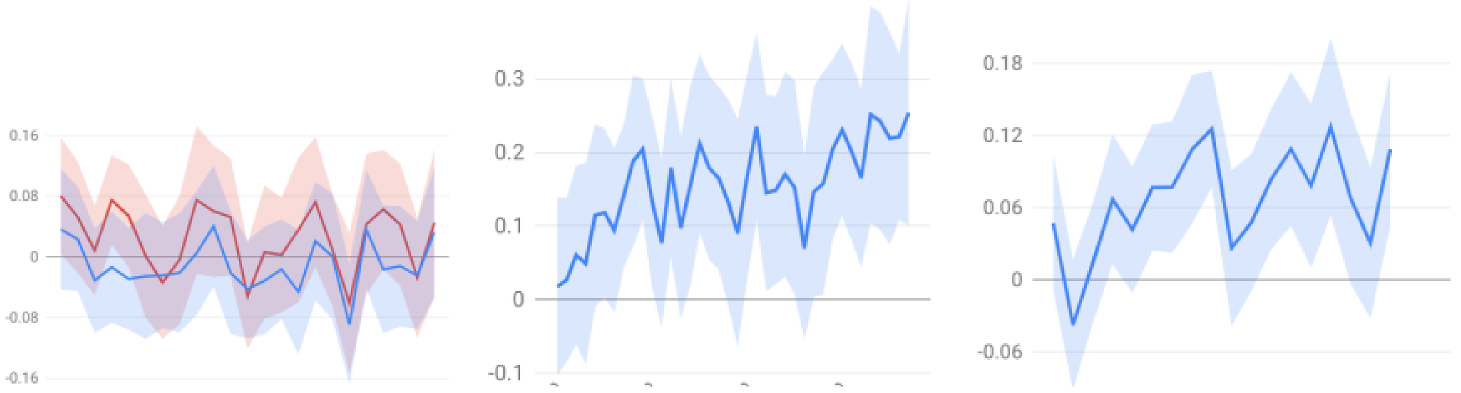
Long vs. short-term evaluation

- Values of User Exploration in Recommender Systems
  https://dl.acm.org/doi/pdf/10.1145/3460231.3474236
  - Reinforcement learning based RS (learning through rewards given for each recommendation)
  - Reward shaping / Intrinsic motivation (improved reward for relevant items from previously unknown interest clusters)

$$R_t(s_t, a_t) = \begin{cases} c \cdot R_t^e(s_t, a_t) & \text{if recommending } a_t \text{ under } s_t \text{ leads to discovery of previously unknown user interests;} \\ R_t^e(s_t, a_t) & \text{otherwise.} \end{cases} \quad (6)$$

Here $c > 1$ is a constant multiplier.

- Promotes serendipity
- How to transfer this for different algorithms?



(a) Entropy Regularization (b) Intrinsic motivation (c) Intrinsic Motiv. + Actionable Repre.

Figure 3: Overall user enjoyment improvement by comparing (a) Entropy regularization vs base REINFORCE; (b) Intrinsic motivation vs base REINFORCE; (c) Intrinsic motivation + Actionable representation vs Intrinsic motivation.