# Joint Modeling of Candidate and Recruiter Preferences for Fair Two-Sided Job Matching

Clara Rus[1(✉)], Masoud Mansoury[3], Andrew Yates[2], and Maarten de Rijke[1]

[1] University of Amsterdam, Amsterdam, The Netherlands
{c.a.rus,m.derijke}@uva.nl
[2] Johns Hopkins University, Baltimore, MD, USA
andrew.yates@jhu.edu
[3] Delft University of Technology, Delft, The Netherlands
m.mansoury@tudelft.nl

**Abstract.** Recommender systems in recruitment platforms involve two active sides, candidates and recruiters, each with distinct goals and preferences. Most recommendation methods address only one side of the problem, leading to potentially ineffective matches. We propose a two-sided fusion framework that jointly models candidate and recruiter preferences to enhance mutual matches between candidates and recruiters. We also propose a personalized two-sided fusion approach to enhance the fairness of job recommendations. Experiments on the XING recruitment dataset show that the proposed approach improves fairness and compatibility, demonstrating the benefits of incorporating two-sided preferences in fairness-aware recommendations.

**Keywords:** Recommendation · Fairness · Recruitment

## 1 Introduction

Recommender systems are widely used across various domains, from entertainment (e.g., music, movies) and practical services (e.g., food, accommodation) to human-centered matching scenarios (e.g., dating, recruitment). Most recommender platforms assume a one-sided model, where users interact with recommended items. In contrast, on two-sided platforms (e.g., dating, recruitment) two active parties interact [19,25,26]. In dating platforms, the goal is to

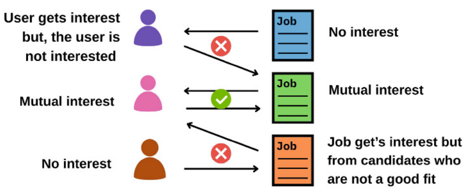


**Fig. 1.** Ignoring dual preferences leads to inefficient matches


R. Campos et al. (Eds.): ECIR 2026, LNCS 16485, pp. 335–351, 2026.
https://doi.org/10.1007/978-3-032-21324-2_27

match people with other people [26], with users belonging to the same pool of users, with similar goals, and engaging with each other's profiles. In contrast, recruitment platforms involve distinct user groups, viz. candidates and recruiters, who interact with different interests: candidates explore and apply to job postings, while recruiters evaluate candidate profiles to fill specific roles. Due to the complex nature of two-sided recruitment platforms [21], prior work mostly considers either the perspective of candidates focusing on a job recommender system [3,12,15,44] or the perspective of recruiters focusing on a candidate recommender system [11,14], ignoring the interplay between them. Such one-sided approaches fail to capture the dependency between the two sides, potentially leading to suboptimal or mismatched recommendations [25], where candidates are shown jobs that are unlikely to result in successful matches, and recruiters receive limited or irrelevant applications. Figure 1 illustrates this situation.

**Beyond Accuracy in Two-Sided Matching.** Some prior work [13,38,43] addresses the two-sided nature of recruitment platforms by learning a joint representation of the candidate-recruiter interest, using textual and interaction data to improve the quality of matches. But as recruitment is a high-stake domain with significant societal implications [1,5,7,29], the focus should go beyond matching quality. Most fairness approaches focus on a one-sided problem, being fair towards the user side [18,27,39] or towards the item side [16,40,41]. Some approaches consider fairness from a two-sided perspective [23,34,36] but they do not consider the two-sided nature of recruitment platforms, where both users and items have preferences, and limited attention/time. Considering the two-sided preferences of a recruitment platform not only enhances engagement and the likelihood of successful matches but also has the potential to improve outcomes for marginalized groups. Promoting candidates/jobs, for the sake of fairness, to recruiters/candidates, who are unlikely to engage with them, whether due to bias or preference, can waste time for both parties and fail to achieve fairness of recruitment outcomes. Recruitment platforms that act as intermediaries may lack information on outcomes beyond profile clicks (e.g., interview offers, hires). While this limits the ability to measure fairness of outcomes, incorporating mutual interest into fairness-aware ranking can serve as a meaningful step towards improving the quality and fairness of matches in recruitment.

**Main Contributions.** We propose a two-sided fusion (TSF) approach for job recommendation. TSF explicitly models the preferences of both candidates and recruiters by fusing the outputs of two independently trained recommender systems: a candidate and a job recommender system. We aim to promote candidate-job pairs that share a mutual interest to the top of the ranking, thereby improving the quality of the matches, and possibly the favorable recruitment outcomes for candidates. Moving beyond quality of matches, we focus on both user (candidate) fairness and item (job) fairness, and propose optimizing the fusion approach under fairness constraints. Code and examples are available on GitHub.[1]

[1] https://github.com/ClaraRus/Joint-Modeling-Job-Matching.

**Main Findings.** We show that TSF improves the fairness, diversity and quality of matches. Depending on the balance between candidate and recruiter signals, TSF can balance item-user fairness. Moreover, we show that by optimizing TSF under fairness constraints, it obtains significant fairness improvements while improving the compatibility between candidates and job offers.

**Theory of Change.** By modeling the two-sided preferences of candidates and recruiters to enhance two-sided fairness, diversity and quality of matches, TSF can improve the recruitment outcomes for marginalized groups. TSF prioritizes recommending jobs for which there is a mutual interest. This relies on having recruiters who showed interest (e.g., click on the profile) in candidates from marginalized communities and assumes that not all recruiters who registered on the platform have a prejudice against candidates from marginalized communities. From conversations with recruitment companies, enhancing diversity and item fairness helps in redistributing applicants to jobs, reducing competitiveness for popular jobs, and improving overall hiring opportunities for candidates, as well as helping the jobs get more exposure and applicants with higher compatibility.

## 2 Related Work

Several works have tackled the two-sided problem in a recruitment setting using the bidirectional signals between candidates and recruiters. Most work focused on learning a joint embedding that captures both the candidate and the recruiter preferences. E.g., BOSS [13] uses textual and interaction data, and models the sequential nature of interactions in a recommendation process (e.g., view, apply, accept), similarly to [10], who model two-sided multi-behavioral sequences to capture dynamic and complementary user preferences. MIRROR [42] models users and the matching process given a multi-view approach (e.g., differentiating between active and inactive users), while BAMBOO [43] models the interactions from two perspectives, user expectation and competitiveness on the recruitment market. [9] propose a reinforcement learning framework for learning the person-job fit, while considering the local market state. [38] propose a dual graph representation learning approach, that uses both textual information from candidate/job profile/description and interaction data, extending the BPR [28] loss to consider the bidirectional preferences.

The above approaches focus on improving the quality of matches between candidates and recruiters. Looking beyond utility-based metrics is an important step. Several works look at fairness from a two-sided perspective in two-sided platforms [23,24,34–37] but they do not model the bidirectional interactions of two-sided platforms. CP-Fair [23] is a two-sided fairness re-ranking method that ensures parity of exposure between item groups and parity of quality of recommendation between user groups. PCT [34] is a two-sided calibration-aware method which expects that the distribution of recommended items to be inline with the user's past interest, and that the item's group exposure to be inline with a target system-level distribution. TFROM [36] is a two-sided individual-based fairness intervention designed to account for both provider fairness and user fairness. TFROM re-ranks items such that the individual quality of recommendation improves, while each item's provider remains within its fair exposure quota. Finally, [32] focus on fairness in a reciprocal recommender system [25,26], proposing a policy for envy-free fairness achieving a good trade-off between fairness and quality of matches.

Previous work focuses on an in-processing approach for enhancing the quality of the matches while requiring interactions beyond clicks on the platform (e.g., interviews, hires, messages), which some platforms may lack. In contrast, we propose a post-processing approach, making it more practical to plug into existing systems, without requiring complex interaction data. Additionally, we go beyond quality of matches and propose a fairness constrained two-sided fusion approach to enhance both the quality of matches and the two-sided fairness.

# 3 Method

## 3.1 Preliminaries and Notation

We consider a recommendation scenario with a set of users $\mathbf{U} = \{u_1, \ldots, u_N\}$ and a set of items $\mathbf{I} = \{i_1, \ldots, i_M\}$. A recommender system $r$ is defined as a function $\mathcal{R}_r : \mathbf{U} \times \mathbf{I} \rightarrow \mathbb{R}$ that assigns a score $\mathcal{R}_r(u_n, i_m)$ to each useritem pair $(u_n, i_m)$. For a user $u_n$, a recommender system $r$ creates a ranking $\pi_r(u_n, k)$ over $k$ items by sorting the items in descending order of the scores. For measuring fairness we write $I^g \subseteq \mathbf{I}$ to denote the set of items belonging to group $g$ and $H^g_{u_n} \subseteq \mathbf{I}$ for the set of items belonging to group $g$ from the user's past interactions.

## 3.2 Two-Sided Fusion

Given the two-sided nature of recruitment platforms, we propose, TSF, a **two-sided fusion** approach modeling both the candidate and the recruiter interest to enhance the quality of matches. Figure 2 shows an overview of the proposed approach. The pipeline is composed of three main steps. (i) Independently training a candidate recommender system and a job recommender system. (ii) Based on the recommendation scores of each, rank the jobs for each candidate. This creates two ranked lists, one generated based on the scores of the job recommender system, and one based on the scores of the candidate recommender system. Finally, (iii) the ranked lists are fused to generate a new recommendation list that prioritizes the mutual interest between candidate and recruiter.

A **job recommender system** ($R_{\text{job}}$) recommends jobs to the candidates registered on a recruitment platform, based on the candidates' interactions. A **candidate recommender system** ($R_{\text{cand}}$) recommends candidates to jobs, based on the recruiters' interactions with candidates profiles given a job description. Therefore, the predicted score $R_{\text{job}}(\text{candidate}, \text{job})$ for a candidate-job pair should reflect the user's interest towards that job offer, while the predicted score $R_{\text{cand}}(\text{job}, \text{candidate})$ for a job-candidate pair should reflect the recruiter's interest towards that candidate. We propose to use the two-sided predicted interest

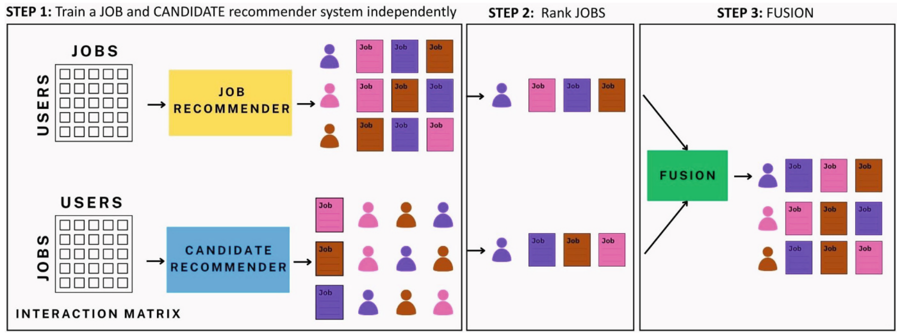


**Fig. 2.** Pipeline of the proposed two-sided fusion method

of a candidate-job relationship to prioritize recommending jobs where the candidate and the recruiter share a mutual interest. For each user $u_n$ we generate two ranked lists: $\pi_{\text{job}}(u_n)$, based on the predicted scores by the job recommender, and $\pi_{\text{cand}}(u_n)$, based on the predicted scores by the candidate recommender. Next, we propose to use a **fusion** method to compute a **fusion score** $f(u_n, i_m)$:

$$f(u_n, i_m, \alpha) = \alpha \mathrm{R}_{\text{job}}(u_n, i_m) + (1 - \alpha)\mathrm{R}_{\text{cand}}(i_m, u_n), \tag{1}$$

where $\alpha$ is a weighting term that represents the importance of candidate side signals vs. recruiter side signals for the final fusion-score. Finally, the fusion score is used to rank the items in descending order of the score, to generate a new ranking of the recommendations: $\pi_{\text{job-cand}}(u_n, k) = \text{argsort}_{i_m \in \mathbf{I}} \left(f(u_n, i_m)\right)_{[1:k]}$.

The fusion function $f$ can be replaced by any existing fusion method. In this work we experimented with both **score based fusion** methods, like SUM [8] (Eq. 1), MIN [8] and MAX [8], and **rank based fusion** methods, like RRF [6], ISR [22] and Borda [2]. Score-based fusion methods aggregate the outputs of multiple recommenders by combining their predicted relevance scores, while rank-based fusion methods, ignore the scores and instead combine the orderings produced by the recommenders, using positional information. This is especially useful when the scores produced by the systems are not available.

### 3.3 Optimized Fusion Method

The fusion function (Eq. 1) assumes a global parameter $\alpha$ that balances the contribution of $R_{\text{job}}$ and $R_{\text{cand}}$ to the final fusion score. This global choice enforces the same balance for all users, regardless of individual preferences, opportunities or fairness considerations towards under-exposed groups of items. To address these limitations, we propose to learn a personalized $\alpha$ parameter per user $u_n$. Then, the fused score is computed as: $f(u_n, i_m) = \alpha_{u_n} s_{\text{job}}(u_n, i_m) + (1 - \alpha_{u_n}) s_{\text{cand}}(i_m, u_n)$, where $\alpha_{u_n}$ is a learned user personalized hyperparameter, balancing the contribution of the candidate-recruiter side. Finally, the personalized fusion score is used to generate a new ranking: $\pi_{\text{job-cand}}(u_n, k, \alpha_{u_n}) =$

$\mathrm{argsort}_{i_m \in \mathbf{I}} \left( f(u_n, i_m, \alpha_{u_n}) \right)_{[1:k]}$. To learn $\alpha_{u_n}$ we propose two optimized fusion approaches: **self-attention fusion (TSF-ATT)** and **fair constrained optimization fusion (TSF-Fair)**.

**Self-attention Fusion.** Using a fixed global parameter $\alpha$ to merge candidate- and recruiter-side signals overlooks user-specific differences in how much each source contributes to relevance. For example, a candidate exploring jobs abroad may benefit more from recruiter-side signals, while another focused locally may rely primarily on past interactions. To capture this heterogeneity, we propose a **self-attention fusion** mechanism that learns personalized fusion weights for each user $u_n$, dynamically balancing candidate- and recruiter-side signals. Similar approaches have been used for combining short- and long-term user preferences in recommendation systems [30,31].

For each user $u_n$, let $R_{\text{job}}(u_n, \cdot) \in \mathbb{R}^M$ and $R_{\text{cand}}(\cdot, u_n) \in \mathbb{R}^M$ denote the predicted relevance vectors from the job and recruiter recommender models. These vectors are stacked as $X_{u_n} = [R_{\text{job}}(u_n, \cdot), R_{\text{cand}}(\cdot, u_n)] \in \mathbb{R}^{2\times M}$ allowing the attention network to process both signals jointly. Each vector $X_{u_n,z}$ ($z \in \{\text{job}, \text{cand}\}$) is passed through a shared feed-forward layer with ReLU activation:

$$h_z = \text{ReLU}(W X_{u_n,z} + b_1), \tag{2}$$

where $W \in \mathbb{R}^{H\times M}$ projects each embedding into a hidden space. The hidden states are then mapped to scalar *attention scores*:

$$w_z = V^\top h_z + b_2, \tag{3}$$

where $V \in \mathbb{R}^H$ is a learnable parameter. The attention scores $[w_{\text{job}}, w_{\text{cand}}]$ are normalized via softmax $[\alpha_{\text{job}}, \alpha_{\text{cand}}] = \text{Softmax}([w_{\text{job}}, w_{\text{cand}}])$ ensuring $\alpha_{\text{job}} + \alpha_{\text{cand}} = 1$. The fused relevance score for user $u_n$ and item $i_m$ is then derived using Eq. 1 where $\alpha_{job}$ and $\alpha_{cand}$ are used as coefficients for $R_{job}$ and $R_{cand}$.

The fusion model is trained end-to-end using implicit feedback data, with the fused scores $f(u_n, \cdot)$ compared to user-item interaction labels via a cross-entropy loss. Parameters are optimized with Adam and weight decay. This adaptive attention mechanism allows the model to personalize fusion: some users rely more on their historical preferences (high $\alpha_{\text{job}}$), while others are better represented by recruiter-side affinities (high $\alpha_{\text{cand}}$). At inference, the fused scores are used for ranking, yielding balanced and context-aware recommendations.

**Fair Constrained Optimization Fusion.** TSF-Fair optimizes the fusion step of TSF by learning a personalized $\alpha_{u_n}$ parameter under fairness constraints. Therefore, TSF-Fair makes use of the two-sided fusion in order to increase both mutual interest and the fairness of the recommendation list towards underexposed groups of items. For each user $u_n$, TSF-Fair learns a personalized fusion parameter $\alpha_{u_n}$ by minimizing a compound objective (Eq. 5) that jointly accounts for multiple fairness and/or preference criteria. We perform a grid search over $\alpha \in [0, 1]$ with a step size of 0.01 to select the value that minimizes the compound loss. Formally, for a given user $u_n$, the optimization problem is defined as:

$$\alpha^*_{u_n} = \arg \min_{\alpha \in [0,1]} \mathcal{L}(u_n, k; \alpha), \tag{4}$$

where $\mathcal{L}$ denotes the compound loss function, defined as:

$$\mathcal{L}(u_n, k, \alpha) = \sum_{m \in \mathcal{M}} \lambda_m \, m\big(\pi_{\text{job-cand}}(u_n, k; \alpha), A\big), \tag{5}$$

where $\mathcal{M}$ is the set of metrics considered during optimization, $\lambda_m$ denotes the relative weight assigned to each metric $m$, and $A$ is the set of sensitive attributes (e.g., country, premium) used to compute the fairness metrics. The fused ranking $\pi_{\text{job-cand}}(u_n, k; \alpha_{u_n})$ is generated by sorting items according to the fusion score function $f(u_n, i_m; \alpha_{u_n})$, defined in Eq. 1, using the learned personalized $\alpha_{u_n}$, which minimizes Eq. 5. The fairness metrics that we consider for optimization are defined in Sect. 4; however, one can make use of alternative fairness metrics. We experiment with $\lambda_m = 1$ but one can set it to different values to balance the trade-off between item-side fairness and user-side fairness.

## 4 Experimental Setup

**Table 1.** Distribution of candidates/jobs and interactions across attribute groups.

| Item | Attribute | Group | # Items | # Interactions |
|---|---|---|---|---|
| Job | Country | DE | 1,073,051 | 314,458 |
| | | non-DE | 204112 | 25,747 |
| | Premium | Yes | 161,060 | 244,378 |
| | | No | 1,116,103 | 95,827 |
| Candidate | Country | DE | 1,082,607 | 2,291 |
| | | non-DE | 209,307 | 201 |
| | Premium | Yes | 275,399 | 684 |
| | | No | 1,016,515 | 1,808 |

**Dataset.** We conduct experiments on the XING dataset, created as part of the RecSys 2017 challenge,[2] which is the only publicly-available recruitment dataset containing two-sided interactions. It contains structured semi-synthetic data collected from the XING platform,[3] a german (DE) recruitment platform, and interaction data from candidates to job openings, and from recruiters to user profiles. Table 1 shows an overview of the data distribution

[2] https://www.recsyschallenge.com/2017/.
[3] https://www.xing.com/.

across candidates/job groups. The available attributes are country of the candidate/job, and whether the candidate/company payed for a premium subscription. The interaction data received by candidates is rather sparse [21]. This makes it difficult to train a candidate recommendation system on the recruiter interaction data. To overcome this problem, in our experiments, we use for training $R_{cand}$, the transpose interaction matrix as a proxy [33] with has an overlap of 55.3% with the real recruiter data.

We address data sparsity and the large scale of the XING dataset by constructing a representative sample of user–item interactions. We first binarize the interaction data: {*click*, *bookmark*, *reply*} to 1 (positive) and the rest to 0 (negative). We extract a *core-20* subset from the positive interactions, by retaining users with at least 20 job interactions and jobs with at least 20 user interactions. From this filtered subset, we randomly sample 10,000 users and include their complete interaction histories to form the dataset used in our experiments. The resulting sample contains 10,000 users with 373,580 positive interactions across 64,063 jobs, corresponding to an interaction density of approximately 0.058%. We split the user profiles into train (80%) for building $R_{\text{job}}$ and $R_{\text{cand}}$, val (10%) for hyperparameter tuning, and test (10%) for evaluation.

**Evaluation Metrics.** As recruitment platforms are inherently multi-stakeholder systems [4,17], fairness should be ensured for both candidates and companies posting job offers. We therefore evaluate our approach in terms of: (i) item-side (job) fairness as group parity between opportunity ratios, (ii) user-side (candidate) as the user preferences with respect to the distribution of the past interactions on particular item groups, (iii) diversity using the gini index, (iv) quality of matches by measuring the compatibility between candidate-job profile, and (v) utility using the NDCG. Item-side fairness is measured in two ways: (i) per user as **RGI** which is defined as:

$$\text{RGI} = \left| \frac{1}{|\mathbf{U}|} \sum_{u \in \mathbf{U}} \frac{\text{Repr}^{g_1}_{\text{user}}(u)}{\%I^{g_1}} - \frac{1}{|\mathbf{U}|} \sum_{u \in \mathbf{U}} \frac{\text{Repr}^{g_2}_{\text{user}}(u)}{\%I^{g_2}} \right|, \tag{6}$$

where $\text{Repr}^{g}_{\text{user}}$ is the proportion of items from group $g$ in the top-$k$ of a user, ensuring fairness in each recommended list and (ii) globally as **RGI(global)** defined:

$$\text{RGI(global)} = \left| \frac{\text{Repr}^{g_1}_{\text{global}}(u)}{\%I^{g}_1} - \frac{\text{Repr}^{g_2}_{\text{global}}(u)}{\%I^{g}_2} \right|, \tag{7}$$

where $\text{Repr}^{g}_{\text{global}}$ is the proportion of items from group $g$ in the top-$k$ across all recommendations lists. User-side fairness is measured as **RUP** which is defined:

$$\text{RUP} = \left| \frac{1}{|\mathbf{U}|} \sum_{u \in \mathbf{U}} \frac{\text{Repr}^{g_1}_{\text{user}}(u)}{\%H^{g_1}_u} - \frac{1}{|\mathbf{U}|} \sum_{u \in \mathbf{U}} \frac{\text{Repr}^{g_2}_{\text{user}}(u)}{\%H^{g_2}_u} \right|, \tag{8}$$

where $\%H^g_u$ is the percentage of the group in the user history of a user. Quality of matches is measured as **compatibility** score computing the % match between the mutual candidates and job features ($F = \{\texttt{industry}, \texttt{discipline}, \texttt{seniority}\}$).

$$\mathrm{Comp}(u, i) \;=\; \frac{1}{|F|} \sum_{f \in F} \mathbb{I}[x_f(u) = x_f(i)]\,, \tag{9}$$

### 4.1 Baselines

For our base recommender systems we use BPR [28], `LibRec-Auto` [20] implementation, to train the job recommender system and the candidate recommender system. We compare our approach to state-of-the art fairness-aware post-processing approaches, including one-sided approaches: **FA*IR** [40] an item-side fairness intervention that aims to maintain the protected group representation above a specified minimum (p=0.6) at every top-$k$, and **User-fairness** [18] is a user-side fairness intervention which aims at balancing the quality of the received recommendations (precision@k) between user-groups, and two-sided approaches: **CP-Fair** [23] a two-sided group fairness re-ranking method, and **TFROM** [36] a two-sided individual-based fairness intervention. We do not compare with existing two-sided approaches for enhancing quality of matches as they are in-processing approaches and require more complex recruiter data which public datasets lack. Since FA*IR does not support multiple attributes, we define the protected group as the most disadvantaged intersectional group.

## 5 Results

### 5.1 Two-Sided Fairness

Table 2 shows an overview of our approach in terms of item (RGI) and user (RUP) fairness in top-10, with lower values indicating an improvement. A lower RGI indicates smaller disparity in the exposure of the groups proportional to their availability, while higher RGI indicates an overrepresentation of one of the groups. Improvements in RGI means that more exposure is given to underrepresented job groups and candidates are being exposed to a more balanced and representative set of job opportunities. A lower RUP indicates that the recommendations are inline with the users' historical preferences towards one attribute. This is relevant especially in the context of the country attribute, while for the premium attribute this is not trivial as the user is not aware of whether the company payed for a premium subscription or not.

**Item Fairness.** As expected, FA*IR obtains the best improvements in terms of item-side fairness, as it is an item-side fairness approach. The user-fairness baseline does not affect item-side fairness metrics, as its optimization objective targets parity in recommendation quality. In our setting, BPR already achieves near parity, resulting in minimal changes to the recommendation list. Out of the two-sided fairness approaches only CP-Fair makes improvements. TFROM prioritizes obtaining similar quality of recommendation across individual users, resulting in little changes on the recommendation list, and thus no changes on the item-side fairness metrics. Our proposed TSF obtains improvements on all item-fairness

**Table 2.** Evaluation in terms of item/user fairness, diversity, utility and quality of matches on the XING dataset at Top-10; "prem." is short for "premium"; **bold** showing best result and underline an improvement

| | Country | | | Premium | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | RGI | | | RGI | | | | |
| Method | RGI↓ | (global)↓ | RUP↓ | RGI↓ | (global)↓ | RUP↓ | Gini↓ | NDCG↑ | Comp↑ |
| BPR | 1.00 | 0.89 | 2.56 | 7.57 | 7.57 | **1.49** | 0.85 | **0.16** | 0.20 |
| FA*IR(country) | **0.34** | **0.34** | 9.68 | 7.51 | 7.51 | 1.50 | 0.85 | **0.16** | 0.20 |
| FA*IR(prem.) | 0.92 | 0.77 | 3.04 | **5.09** | 5.08 | 4.23 | 0.85 | 0.15 | 0.20 |
| FA*IR(country,prem.) | 0.50 | 0.49 | 10.54 | 5.84 | 5.84 | 2.96 | 0.85 | **0.16** | 0.20 |
| userfairness(country) | 1.00 | 0.89 | 2.57 | 7.57 | 7.57 | **1.49** | 0.85 | **0.16** | 0.20 |
| userfairness(prem.) | 1.00 | 0.89 | 2.57 | 7.57 | 7.57 | **1.49** | 0.85 | **0.16** | 0.20 |
| CP-Fair(country) | 1.00 | 0.88 | 2.57 | 7.58 | 7.57 | **1.49** | 0.85 | **0.16** | 0.20 |
| CP-Fair(prem.) | 1.00 | 0.88 | 2.59 | 7.54 | 7.51 | **1.49** | 0.85 | **0.16** | 0.20 |
| TFROM(country) | 1.00 | 0.89 | 2.56 | 7.57 | 7.57 | **1.49** | 0.85 | **0.16** | 0.20 |
| TFROM(prem.) | 1.00 | 0.89 | 2.56 | 7.57 | 7.57 | **1.49** | 0.85 | **0.16** | 0.20 |
| TFROM(country,prem.) | 1.00 | 0.89 | 2.56 | 7.57 | 7.57 | **1.49** | 0.85 | **0.16** | 0.20 |
| TSF-ATT | 1.00 | 0.88 | 2.55 | 7.28 | 7.28 | 1.70 | 0.83 | 0.15 | 0.20 |
| TSF-Fair(RGI(country)) | 0.75 | 0.57 | 3.62 | 6.77 | 6.77 | 1.61 | 0.69 | 0.10 | **0.24** |
| TSF-Fair(RGI(prem.)) | 1.07 | 0.77 | 1.86 | 6.37 | 6.36 | 1.60 | **0.64** | 0.12 | **0.25** |
| TSF-Fair(RUP(country)) | 1.00 | 0.75 | **1.69** | 6.63 | 6.61 | 1.62 | 0.65 | 0.12 | **0.25** |

metrics for both the country and the premium attribute, depending on the type of optimization, except for RGI(country) for which only TSF-Fair(RGI(country)) obtains improvements. Overall, TSF obtains better improvements than the user-side and two-sided fairness approaches, but lower than of the item-side approach. When looking at the diversity, only TSF makes improvements. As TSF incorporates recruiter-side signals, if a candidate receives low interest from recruiters associated with one group of jobs but high interest from recruiters associated with another, the fusion will promote jobs from the latter group, thus enhancing both the item-side fairness and the diversity.

**User Fairness.** Respecting user's preferences with respect to the premium attribute is not trivial, since candidates are not aware of whether a job is premium. In contrast, taking into account the user's preference for a specific country is more important, as users who are not willing to relocate will not be interested in jobs abroad. None of the baselines show improvements in terms of respecting the user preferences. FA*IR, the item-side approach, shows the highest increase in the discrepancy with the user preference, while user-side and two-sided approaches show only a slight increase, showing that they are more inline with the users preference even though their user-side optimization objective is based on quality of recommendations. In contrast, TSF shows improvements for RUP(country).

**TSF-Fair.** We propose learning a personalized $\alpha_u$ hyperparameter given some fairness constraints (TSF-Fair). TSF-Fair outperforms TSF-ATT on all metrics, achieving the largest gains for the metric it was optimized for. Notably, on RUP(country) TSF-Fair obtains the best outcomes. There seem to be trade-off in optimizing for RGI(country) and RUP(country); optimizing for one comes at the trade-off of the other. All TSF-Fair variants improve RGI(global) for both attributes. Interestingly, optimizing for RGI(country) improves RGI(premium), while optimizing for RGI(premium) does not improve the RGI(country). Comparing TSF-Fair with the use of a global $\alpha$ (Fig. 4 TSF-SUM), we observe that TSF-Fair outperforms the global-$\alpha$ approach. Moreover, TSF-Fair makes significant improvements in terms of diversity, obtaining the best improvements.

**Utility.** In contrast to most baselines, TSF shows a decrease in utility. This is expected, as TSF de-prioritizes offers reflecting only the candidates interest, however, this can benefit the users by prioritizing offers from interested recruiters.

Overall, TSF, our proposed two-sided fusion method, can obtain consistent improvements across item-side fairness metrics for both the country and the premium attribute, while obtaining improvements in terms of user preferences, keeping a good balance between item-user fairness, but with a trade-off on utility. Additionally it makes significant improvements in terms of diversity.

## 5.2 Candidate-Job Compatibility

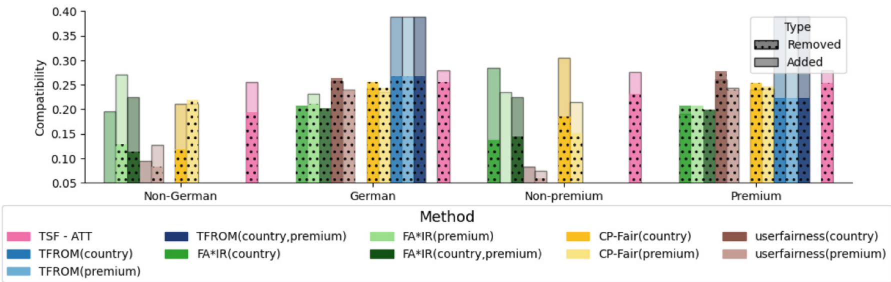


**Fig. 3.** Compatibility@10 of added/removed items in comparison with BPR

To measure the quality of matches we compute a compatibility score between the candidate profile and job offer. We argue that by increasing the compatibility, we help increase the mutual interest, as candidates are more likely to interact with jobs aligning with their skills, and recruiters with candidates whose skills align with the job offer. Table 2 shows an overview of the compatibility scores (Comp). It can be observed that none of the baselines affect the overall compatibility. In contrast, TSF-Fair improves the compatibility, as we take into account both the recruiter and the candidate signals, boosting job offers with mutual interest, and therefore, mutual compatibility. This shows that the joint modeling of candidate-recruiter signals, under fairness goals can obtain significant fairness improvements, while improving the quality of matches between candidates and recruiters. This could potentially benefit the stakeholders of a recruitment platform: (i) the candidate – as they are more likely to be recommended jobs for which they are a good fit and for which there is a mutual interest, (ii) the recruiter – as they get applications from fitting candidates, and more exposure, and (iii) the platform as it gains more engagement.

Next, we analyze the compatibility of added vs removed items (Fig. 3). Added items appear in the new recommendation list but not in the BPR's, while removed items are those removed from BPR. Ideally, added items should have higher compatibility than removed ones across all groups. This gives insight into whether the candidates are losing more than winning in terms of compatibility. As expected, FA*IR removes jobs from the disadvantaged groups and adds jobs from the advantaged ones, with overall added jobs having higher compatibility. However, FA*IR(country) affects the premium attribute by removing some items with higher compatibility than the ones added. User-fairness adds underrepresented jobs with higher compatibility than the ones removed, but for user-fairness(country) the removed premium ones have a higher compatibility than the added ones. For CP-Fair the compatibility of items added of the disadvantaged groups is higher than of the removed ones, while for the the advantaged groups the compatibility of the removed ones is higher than of the added. CP-Fair(premium) removes some non-german jobs with higher compatibility than the ones added. All TFROM approaches increase the compatibility of added items much more than the other approaches, but only for the advantaged groups. There are no changes in the recommendation lists with respect to the disadvantaged groups. Overall, TSF-ATT consistently adds items with higher compatibility than the ones removed across all item groups.

### 5.3 Balancing Recruiter and Candidate Signals

In this section we discuss how the $\alpha$ hyperparameter of balancing the candidate-recruiter signals in the fusion impacts the two-sided fairness, as well as the compatibility and utility, over various fusion strategies (Fig. 4).

TSF with a global $\alpha$ shows similar trends over various fusion strategies, except for TSF-MAX/MIN, which takes the fusion score to be the max/min score of one of the lists. TSF-MAX favors jobs highly scored by either side, while TSF-MIN prioritizes mutual agreement on the scores (lower score discrepancy). Lowering $\alpha$ the diversity of the jobs recommended improves, as the gini index is lower. Analyzing the job groups, we observed that lowering $\alpha$ increases the overall number of non-premium and non-german jobs, groups which are overall underrepresented in the recommendations. This improves the global item-side fairness metric for both country and premium. For premium we observed same trend between RGI per user and global, while the RUP is worse, but respecting the user preferences for the premium attribute is not trivial. For the country attribute, we can observe a trade-off between item-side fairness per user (RGI) and respecting

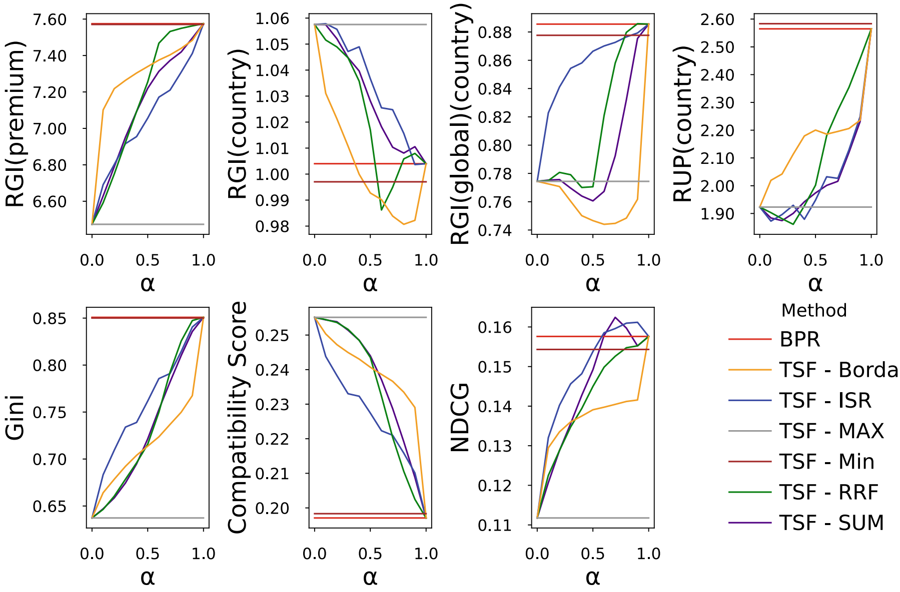


**Fig. 4.** Effect of the fusion parameter $\alpha$ on two-sided fusion. Lower $\alpha$ gives more weight to recruiter signals, while higher $\alpha$ to candidate signals

the user preferences (RUP). It seems that increasing the number of underrepresented jobs in the recommendations, helps in respecting the user preferences. By increasing the number of non-german jobs in the recommendations, candidates with an interest in non-german jobs get their preferences respected, but this comes at the cost of RGI as respecting the proportionality of the item groups in each recommendation list is not possible. Lowering $\alpha$ favors the recruiter signals; if a candidate receives low interest from recruiters associated with one group but high interest from recruiters associated with another, the fusion will promote jobs from the latter group, thus depending on the interest and distribution of recruiters, this enhances diversity leading to improvements in fairness. Compatibility increases when lowering $\alpha$ as it emphasizes recruiter signals, which better reflect candidatejob fit; this reduces utility by de-prioritizing user interest. In contrast, higher $\alpha$ favors user preferences, while lowering compatibility, as user's interest might not necessary reflect a good profile match.

## 6 Conclusion

We have modeled recruitment platforms as two-sided platforms with two active and distinct pools of users with independent goals. Candidates aim to find a job given their personal preference and goals, while recruiters aim to find the best candidate for the job description. We propose TSF, a two-sided fusion approach, which models the two-sided preferences to enhance the quality of the matches.

We go beyond quality of matches, and show that TSF can be used for enhancing fairness and diversity, while improving or maintaining the compatibility. TSF can obtain significant fairness improvements, while improving compatibility. TSF can improve recruitment outcomes for candidates by recommending jobs with higher mutual compatibility and by increasing the diversity of the jobs recommended, which in turn helps distribute applications more evenly and reduces competitiveness for popular jobs. Due to the lack of public data, this research is limited to only one recruitment dataset, for which the recruiter interactions were very sparse, resulting in using the transpose user-item interaction matrix instead. Future work could focus on designing more comprehensive recruitment datasets, or simulators specifically designed for recruitment, especially for the recruiter signals, to help understand under which conditions the proposed method fails, and whether unwanted biases are introduced.

**Acknowledgments.** We thank our reviewers for the feedback. This research was supported by the EU's Horizon Europe program (No. 101070212 and 101201510), and by NWO (VI.Vidi.223.166, 024.004.022, NWA.1389.20.183, and KICH3.LTP.20.006). All content represents the opinion of the authors, which is not necessarily shared or endorsed by their respective employers and/or sponsors.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Ajayi, F.A., Udeh, C.A.: Innovative recruitment strategies in the IT sector: a review of successes and failures. Magna Sci. Adv. Res. Rev. **10**(2), 150–164 (2024)
2. Aslam, J.A., Montague, M.H.: Models for metasearch. In: SIGIR 2001: Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 275–284. September 9–13, 2001, New Orleans, Louisiana, USA, ACM (2001)
3. Azri, M.H.F.B., Haw, S., Ng, K.W., Saad, M.Y.B.M.: Context-aware job recommender system. JOIV: Int. J. Inf. Vis. **9**(2), 877–886 (2025)
4. Beretta, A., et al.: Requirements of explainable AI in algorithmic hiring. In: Proceedings of the AIMMES 2024 Workshop on AI Bias: Measurements, Mitigation, Explanation Strategies, co-located with the EU Fairness Cluster Conference 2024, Amsterdam, The Netherlands (2024)
5. Chen, Z.: Ethics and discrimination in artificial intelligence-enabled recruitment practices. Humanit. Soc. Sci. Commun. **10**(1), 1–12 (2023)
6. Cormack, G.V., Clarke, C.L.A., Büttcher, S.: Reciprocal rank fusion outperforms condorcet and individual rank learning methods. In: SIGIR, pp. 758–759. ACM (2009)
7. Fabris, A., et al.: Fairness and bias in algorithmic hiring: a multidisciplinary survey. ACM Trans. Intell. Syst. Technol. **16**(1), 1–54 (2025)
8. Fox, E.A., Shaw, J.A.: Combination of multiple searches. In: TREC, NIST Special Publication, vol. 500-215, pp. 243–252. National Institute of Standards and Technology (NIST) (1993)

9. Fu, B., et al.: Market-aware dynamic person-job fit with hierarchical reinforcement learning. In: International Conference on Database Systems for Advanced Applications, pp. 697–705. Springer, Cham (2022)
10. Fu, B., Liu, H., Zhu, Y., Song, Y., Zhang, T., Wu, Z.: Beyond matching: modeling two-sided multi-behavioral sequences for dynamic person-job fit. In: Jensen, C.S., et al. (eds.) International Conference on Database Systems for Advanced Applications, pp. 359–375. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-73197-7_24
11. Geyik, S.C., Ambler, S., Kenthapadi, K.: Fairness-aware ranking in search & recommendation systems with application to linkedin talent search. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2221–2231 (2019)
12. Giabelli, A., Malandri, L., Mercorio, F., Mezzanzanica, M., Seveso, A.: Skills2job: a recommender system that encodes job offer embeddings on graph databases. Appl. Soft Comput. **101**, 107049 (2021)
13. Hu, X., Cheng, Y., Zheng, Z., Wang, Y., Chi, X., Zhu, H.: Boss: A bilateral occupational-suitability-aware recommender system for online recruitment. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 4146–4155 (2023)
14. Huang, Y., Liu, D.R., Lee, S.J.: Talent recommendation based on attentive deep neural network and implicit relationships of resumes. Inf. Process. Manage. **60**, 103357 (2023)
15. Hulbatte, S.: Enhanced job recommendation system. Int. J. Res. Appl. Sci. Eng. Technol. **7**(V), 3556–3563 (2019)
16. Jiang, M., et al.: Item-side fairness of large language model-based recommendation system. In: Proceedings of the ACM Web Conference 2024, pp. 4717–4726 (2024)
17. Kaya, M., Bogers, T.: Mapping stakeholder needs to multi-sided fairness in candidate recommendation for algorithmic hiring. In: Proceedings of the Nineteenth ACM Conference on Recommender Systems, pp. 257–267 (2025)
18. Li, Y., Chen, H., Fu, Z., Ge, Y., Zhang, Y.: User-oriented fairness in recommendation. In: Proceedings of the Web Conference 2021, pp. 624–632 (2021)
19. Malinowski, J., Keim, T., Wendt, O., Weitzel, T.: Matching people and jobs: a bilateral recommendation approach. In: Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06), vol. 6, pp. 137c–137c (2006)
20. Mansoury, M., Burke, R., Ordonez-Gauger, A., Sepulveda, X.: Automating recommender systems experimentation with librec-auto. In: Proceedings of the 12th ACM Conference on Recommender Systems, pp. 500–501 (2018)
21. Mashayekhi, Y., Li, N., Kang, B., Lijffijt, J., De Bie, T.: A challenge-based survey of e-recruitment recommendation systems. ACM Comput. Surv. **56**(10), 1–33 (2024)
22. Mourão, A., Martins, F., Magalhães, J.: Multimodal medical information retrieval with unsupervised rank fusion. Comput. Med. Imaging Graph. **39**, 35–45 (2015)
23. Naghiaei, M., Rahmani, H.A., Deldjoo, Y.: CPFair: personalized consumer and producer fairness re-ranking for recommender systems. In: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 770–779 (2022)
24. Patro, G.K., Biswas, A., Ganguly, N., Gummadi, K.P., Chakraborty, A.: Fairrec: two-sided fairness for personalized recommendations in two-sided platforms. In: Proceedings of the Web Conference 2020, pp. 1194–1204 (2020)

25. Pizzato, L., Rej, T., Akehurst, J., Koprinska, I., Yacef, K., Kay, J.: Recommending people to people: the nature of reciprocal recommenders with a case study in online dating. User Model. User-Adap. Inter. **23**(5), 447–488 (2013)
26. Pizzato, L., Rej, T., Chung, T., Koprinska, I., Kay, J.: RECON: a reciprocal recommender for online dating. In: Proceedings of the Fourth ACM Conference on Recommender Systems, pp. 207–214 (2010)
27. Rastegarpanah, B., Gummadi, K.P., Crovella, M.: Fighting fire with fire: using antidote data to improve polarization and fairness of recommender systems. In: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, pp. 231–239 (2019)
28. Rendle, S., Freudenthaler, C., Gantner, Z., Schmidt-Thieme, L.: BPR: Bayesian personalized ranking from implicit feedback. In: Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, pp. 452–461 (2009)
29. Rus, C., Luppes, J., Oosterhuis, H., Schoenmacker, G.H.: Closing the gender wage gap: adversarial fairness in job recommendation. arXiv preprint arXiv:2209.09592 (2022)
30. Sabouri, M., Mansoury, M., Lin, K., Mobasher, B.: Temporal user profiling with LLMs: Balancing short-term and long-term preferences for recommendations. arXiv preprint arXiv:2508.08454 (2025a)
31. Sabouri, M., Mansoury, M., Lin, K., Mobasher, B.: Towards explainable temporal user profiling with llms. In: Adjunct Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization, pp. 219–227 (2025b)
32. Tomita, Y., Yokoyama, T.: Fair reciprocal recommendation in matching markets. In: Proceedings of the 18th ACM Conference on Recommender Systems, pp. 209–218 (2024)
33. Vargas, S., Castells, P.: Improving sales diversity by recommending users to items. In: Proceedings of the 8th ACM Conference on Recommender systems, pp. 145–152 (2014)
34. Wang, C., et al.: Two-sided calibration for quality-aware responsible recommendation. In: Proceedings of the 17th ACM Conference on Recommender Systems, pp. 223–233 (2023)
35. Wu, G., et al.: Fairsort: learning to fair rank for personalized recommendations in two-sided platforms. IEEE Trans. Knowl. Data Eng. **37**(2), 641–654 (2024)
36. Wu, Y., Cao, J., Xu, G., Tan, Y.: TFROM: a two-sided fairness-aware recommendation model for both customers and providers. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1013–1022 (2021)
37. Xu, L., Lin, Z., Wang, J., Chen, S., Zhao, W.X., Wen, J.R.: Promoting two-sided fairness with adaptive weights for providers and customers in recommendation. In: Proceedings of the 18th ACM Conference on Recommender Systems, pp. 918–923 (2024)
38. Yang, C., Hou, Y., Song, Y., Zhang, T., Wen, J.R., Zhao, W.X.: Modeling two-way selection preference for person-job fit. In: Proceedings of the 16th ACM Conference on Recommender Systems, pp. 102–112 (2022)
39. Yao, S., Huang, B.: Beyond parity: fairness objectives for collaborative filtering. In: Advances in Neural Information Processing Systems, vol. 30, pp. 2925–2934 (2017)
40. Zehlike, M., Bonchi, F., Castillo, C., Hajian, S., Megahed, M., Baeza-Yates, R.: FA*IR: a fair top-$k$ ranking algorithm. In: Proceedings of the 2017 ACM Conference on Information and Knowledge Management, pp. 1569–1578. ACM (2017)

41. Zehlike, M., Castillo, C.: Reducing disparate exposure in ranking: a learning to rank approach. In: Proceedings of the Web Conference 2020, pp. 2849–2855 (2020)
42. Zheng, Z., Hu, X., Gao, S., Zhu, H., Xiong, H.: Mirror: A multi-view reciprocal recommender system for online recruitment. In: Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 543–552 (2024a)
43. Zheng, Z., et al.: Bilateral multi-behavior modeling for reciprocal recommendation in online recruitment. IEEE Trans. Knowl. Data Eng. **36**(11), 5681–5694 (2024)
44. Zou, Z., Huspi, S.H., Nuar, A.N.A.: A review on job recommendation system. J. Adv. Res. Appl. Sci. Eng. Technol. **41**(2), 113–124 (2024)