Fairness-Aware Group Recommendation with Pareto-Efficiency

Lin Xiao
Tsinghua University
Institute of Interdisciplinary Information Sciences,
jackielinxiao@gmail.com

Zhang Min*
Tsinghua University
Department of Computer Science and Technology,
z-m@tsinghua.edu.cn

Zhang Yongfeng
University of Massachusetts Amherst
College of Information and Computer Science,
yongfeng@cs.umass.edu

Gu Zhaoquan
Guangzhou University, and The University of HongKong
Department of Computer Science,
demin456@gmail.com

Liu Yiqun
Tsinghua University
Department of Computer Science and Technology,
yiqunliu@tsinghua.edu.cn

Ma Shaoping
Tsinghua University
Department of Computer Science and Technology,
msp@tsinghua.edu.cn

ABSTRACT

Group recommendation has attracted significant research efforts for its importance in benefiting a group of users. This paper investigates the Group Recommendation problem from a novel aspect, which tries to maximize the satisfaction of each group member while minimizing the unfairness between them. In this work, we present several semantics of the individual utility and propose two concepts of social welfare and fairness for modeling the overall utilities and the balance between group members. We formulate the problem as a multiple objective optimization problem and show that it is NP-Hard in different semantics. Given the multiple-objective nature of fairness-aware group recommendation problem, we provide an optimization framework for fairness-aware group recommendation from the perspective of Pareto Efficiency. We conduct extensive experiments on real-world datasets and evaluate our algorithm in terms of standard accuracy metrics. The results indicate that our algorithm achieves superior performances and considering fairness in group recommendation can enhance the recommendation accuracy.

1 INTRODUCTION

Group recommendation is to recommend items to groups of users whose preferences can be different from each other. The applications of group recommendation are frequently seen in real life and on the web, for example, families have to decide which TV program to watch; Social network websites [31] encourage users to build up social groups and share contents within groups; and E-Commerce websites like Groupon recommend coupons to groups of users for sales promotion. Other scenarios include music/movie group recommendation [7], and restaurant recommendations to groups [19].

The research of group recommendation was first seen in [6], which introduces a group recommender system called Polylens to recommend movies to groups of users. Previous studies on group recommendation deal with the problem on two kinds of groups: persistent and ephemeral groups [22]. The persistent groups refer to groups with consistent structures and historical records of interactions between groups and items [30] [13] [28]; while ephemeral groups can be formed ad hoc and the users may just constitute the groups for the first time [2] [39] [22]. For recommendation to persistent groups, the group can be seen as a virtual user and the personalized recommendation algorithms can be applied since the interactions between groups and items are available. But for recommendations to ephemeral groups, the historical interactions do not exist and the recommendation can only be generated from the aggregation of the individual preferences. In this paper, we concentrate on the general scenario of making recommendations to ephemeral groups.

Previous studies on group recommendation focus on recommending one item at a time. Various preference aggregation functions are applied to find a consensus between the users on a single item. Some important semantics have been proposed for the evaluation [1], [31] [26]. In this way, the group recommendation algorithms evaluate how an item satisfies a group of users and select the Top K items which satisfy the group most as recommendation. However, as pointed out in [21], the best item according to this function may still leave some users feeling dissatisfied and slighted.

In this paper, we look at the problem from a different perspective of user utility in the group recommendation. The user utility is determined by how relevant the recommended items are to the user. Therefore only when the K recommended items are decided can we know the utility of the user achieved from the items. We propose to evaluate the user satisfaction with utility function and further consider the overall social welfare (the sum of user utilities inside the

*Corresponding Author

RecSys'17, August 27–31, 2017, Como, Italy.
 ISBN 978-1-4503-4652-8/17/08... $15.00.
DOI: http://dx.doi.org/10.1145/3109859.3109887

group) and fairness (the balance of user utilities inside the group) to evaluate the quality of group recommendation.

Evaluating user satisfaction from each individual's perspective can be found in some studies [36] [21]. This modeling makes it possible to explicitly evaluate the imbalances between users in terms of their satisfactions achieved from the group recommendation.

More specifically, we model the individual satisfaction about the recommendation by considering the relevance of each recommended item to the user. Meanwhile, we adopt the semantics originating from Social Choice Theory [35] [3] and fairness measure [14] to combine the individual satisfactions as the fairness of group recommendation. We formulate the fairness-aware group recommendation problem as a multiple objective optimization problem and provide a theoretical analysis to show that it is NP-Hard. We further provide a multiple objective optimization framework which preserves Pareto efficiency between multiple objectives.

The remaining of the work is organized as follows: the next section introduces some important related work about group recommendation and multiple objective optimization; Section 3 presents the detailed modeling of user utilities, social welfare and fairness, then we introduce the problem formulation of fairness-aware group recommendation; We propose Pareto-effcient algorithms to solve the problem with scarlarization approaches in Section 4; We conduct extensive experiments on real-world datasets with different types of groups and evaluate the recommendation quality with typical accuracy metrics in Section 5. Finally, we get to the conclusion in Section 6.

2 RELATED WORK

In this section, we introduce some important studies about group recommendation and multiple objective optimization.

2.1 Group Recommendation

The studies of group recommendation can be categorized into two aspects based on the types of groups [22]. Groups with consistent structures and historical interactions are often referred as persistent groups while the groups formed by users ad-hoc are referred as ephemeral groups. Making recommendations to persistent groups is closer to personalized recommendation, considering that each group can be seen as a user and personalized recommendation techniques can be applied [30] [13] [28]. In our work, we focus on ephemeral group recommendations.

The recommendation to ephemeral groups is much more difficult since there exist no interactions between groups and items thus the group can not be seen as a virtual user. Existing studies on ephemeral group recommendations focus on preference aggregation methods that aggregate the individual preferences of group members as a group preference. The collaborative filtering approaches are utilized to generate the individual preferences first and various preference aggregation functions are used to aggregate the individual preferences into group preferences.

The combination of Collaborative Filtering approaches and group recommendation is reflected in some previous studies: [2], [40], [20] and [37]. Various aggregation functions have been proposed for prefrence aggregation, most of them come from the social choice theory [3]. The important semantics are Average and Least Misery, which are widely adopted in group recommendation studies [1], [2], [29], [32] and [22]. The similarity between our work and previous studies is that both of them compute the individual preferences first. The difference is that our work considers the satisfactions of users from the perspective of each individual and tries to maximize the overall satisfaction while maintaining fairness between them. While existing studies model the relevance of a single item to the group and fairness is not explicitly modeled.

It is assumed that user influences exist in the group decision process, the studies concerning with user influences include [17], [34], [23], [39] and [22]. Some probabilistic models have been proposed to capture the user preferences and their influences in the group. The basic assumption of these studies is that users should be treated differently in group recommendation and the concept of influence is proposed to capture the heterogeneity. In our work, users also have to make compromises with each other in order to achieve better fairness in group recommendation, which is similar to heterogeneous influences in spirit (considering that treating users the same may lead to suboptimal solutions in terms of fairness). Some studies incorporate social relationships into the group recommendation, including [32], [10] and [38]. Social relationships can be useful in enriching the user preferences [32] and some studies incorporates both social and content interests in group recommender systems to enhance the recommendation quality.

There are few studies on the fairness issues in group recommendation. Some previous studies concern with this problem from the perspective from game theory and voting theory: [12], [18], [4], [5] and [16]. Some group recommendation studies treat the group decision process as non-cooperative games and try to find equilibria of the games as feasible recommendations. Some other studies treat the group decision process as a voting campaign and use voting mechanisms to find a proper recommendation through a voting process. However, these approaches do not explicitly consider fairness in the recommendation modeling and the trade-off between overall satisfaction and fairness is unclear.

A similar study on fairness in group recommendation is [21], where user fairness is modeled as how satisfied the user is with the group recommendation. [21] tries to maximize the predefined fairness so that the number of users who find the recommendation is fair is maximized. In this paper, we look at the fairness from a different perspective, we model the balance between user utilities with the group recommendation as fairness and hope to find a proper trade-off between the overall user utilities (social welfare)and the differences between them (fairness).

There are several differences between this work and [21]: First, the fairness semantics are different in two studies. In our work, we first define the individual utility given a recommendation to the group and then treat the imbalances between group members' utilities as the metric of fairness. We argue that the fairness proposed in [21] sometimes may cause extreme unfairness to the users. Consider the the case when a group of ten users are recommended with an item, eight of them are satisfied with the item while two of them hate the item. According to [21], the fairness is 0.8 which means quite a majority of group members find the recommendation "fair". However the remaining two users who dislike the item are ignored in this semantic. In our proposal, the fairness is relatively low since two users suffer from the recommendation severely (both least misery and min-max fairness care about the least satisfied user). Second, [21] concerns with maximization of fairness in group recommendation while our work considers the group recommendation from a multi-objective perspective (optimizing fairness and social welfare simultaneously). Third, [21] evaluates the performances of group recommendation with self-defined metrics while this work evaluates the quality of group recommendation with widely-used accuracy metrics. Moreover, this work provides a general framework for defining fairness with individual utility and an optimizing framework with Pareto efficiency. Therefore the similar procedure can be extended to other fairness semantics in group recommendation.

2.2 Multi-Objective Optimization

The studies on multi-objective optimization are rich and various approaches have been proposed [9]. One important feature of multiple objective optimization is that (usually) there does not exist a solution that satisfies all the objectives simultaneously. In that case, the objective functions are conflicting, and there exist a (possibly infinite) number of Pareto optimal solutions. A solution is called Pareto efficient, if none of the objective functions can be improved without degrading some of the other objectives. Notice that there are multiple solutions that satisfy the Pareto Optimality, the set of all Pareto optimal solutions is named as Pareto Front.

Various ways have been proposed to solve the Pareto Front, including A Priori approaches and A posteriori methods. The first approache aims to scalarize the objectives into a single objective function; while the second approach uses evolutionary algorithms [41] to update the solutions by iterations.

Some studies have considered multiple objectives in personalized recommendation tasks, including [27], [15], [25]. In [25], multiple objectives including accuracy, diversity and novelty are simultaneously considered and a pareto front is found to satisfy the mentioned objectives. There are few studies on optimizing multiple objectives in group recommendation and we are among the first to optimize both overall satisfaction and fairness in group recommendation through a multiple-objective optimization framework.

3 FAIRNESS-AWARE GROUP RECOMMENDATION MODELING

3.1 Individual Utility Modeling

In recommender systems, the preferences of users are learnt from their historical interactions with items. Given a new item, the relevance of the item to a user can be predicted with Collaborative Filtering methods [33] [24]. In the group recommendation process, we first compute the relevances of candidate items to each user. In this way, we get a full user-item interaction matrix where each entry (between user u and item i) reflects the pairwise relevance as $rel(u,i) \in [rel_{min}, rel_{max}]$ where rel_{max} and rel_{min} represent the ranges of relevance values.

The utility function $U(u,I)$ of user u given the recommendation I is a function of the relevances of recommended items $rel(u,i)$, $\forall u, i \in I$ to the user:

Definition 3.1. **Individual Utility:** The individual utility of user u in group G when a set of items I ($|I| = K$) are recommended to the group, is a function $U(u,I) : U \times I \rightarrow [0,1]$ of the relevances $rel(u,i)$ where $i \in I$. Some semantics are presented as follows:

(1) Average: $U(u,I) = \frac{1}{K \times rel_{max}} \sum_{i \in I} rel(u,i)$
(2) Proportionality: $U(u,I) = \frac{\sum_{i \in I} rel(u,i)}{\sum_{i \in I(u,K)} rel(u,i)}$

where $I(u,K)$ denotes the set of items which are among the top-K favourite items of user u. Notice that the relevance of an item to a user can be binary or fractional. When the relevance is 1 for an item from the top-K favourite items, the utility following proportionality semantic becomes the ratio of top-K favorite items of the user in the group recommendation list.

3.2 Social Welfare and Fairness Modeling

For evaluating the overall satisfaction of users about group recommendation quality, we aggregate all the individual utilities as social welfare and further consider fairness as the extent of imbalance between their individual utilities.

Definition 3.2. **Social Welfare:** The Social Welfare (denoted as $SW(g,I)$) is the overall utility of all users inside the group g given a group recommendation I:

$$SW(g,I) = \frac{1}{|g|} \sum_{u \in g} U(u,I), \forall g, I$$

The fairness depicts how imbalanced the users are satisfied with the recommendation. Therefore it should reflect the comparison between the utilities of users inside the group.

Definition 3.3. **Fairness:** Given a group recommendation I to group g, the fairness (denoted as $F(g,I)$) is a function of $U(u,I), \forall u \in g$, $\forall I$. Some semantics are summarized as follows:

$$Least\ Misery:\ F_{LM}(g, I) = \min\{U(u, I), \forall u \in g\}$$
$$Variance:\ F_{Var}(g, I) = 1 - Var(\{U(u, I), \forall u \in g\})$$
$$Jain's\ Fairness:\ F_J(g, I) = \frac{(\sum_{u \in g} U(u, I))^2}{|U| \cdot \sum_{u \in g} U(u, I)^2}$$
$$Min - Max\ Ratio:\ F_M(g, I) = \frac{\min\{U(u, I), \forall u \in g\}}{\max\{U(u, I), \forall u \in g\}}$$

Variance Fairness and Jain's Fairness [14] encourage the group members to achieve close utilities between each other; while Least Misery Fairness and Min-Max Ratio emphasise the gap between the least and highest utilities of group members. Despite the differences of the Fairness in definitions, the consensus intuition of these metrics is to minimize the difference between the individual utilities of group members. We formulate the fairness-aware group recommendation problem as follows:

Problem 1. *Given a group of users U and a set of items $\tilde{I}$, the group recommendation aims to recommend a set $I \in \tilde{I}$ of K items to maximize the social welfare $SW(U, I)$ and fairness $F(U, I)$.*

However, it is difficult to optimize the fairness metrics in group recommendations. We analyze the computational complexity of optimizing the fairness metric and present the results here:

Theorem 3.4. *The Least Misery fairness and Min-Max Ratio fairness maximization problems are NP-Hard.*

Proof. Without loss of generality, we consider a restricted version of the problem where the matrix consists of integer 0s or 1s and has even number of columns. Consider the decision version of the fairness optimization problem: is there a recommendation I of K items for the group g so that $F_{LM}(g, I) \geq \frac{1}{K}$, $F_{Var}(g, I) \geq \frac{1}{K}$? This is equivalent to the problem: **given a matrix A whose entry is either 0 or 1 and the sum of each row is K, is there a selection of K columns such that the sum of each row with selected columns is greater than 1?**

We reduce the problem to 3SAT problem, which is a famous NP-Hard problem. The 3SAT problem is given arbitrary formulas, determining the satisfiability of a formula in conjunctive normal form where each clause is limited to three literals. For any given instance $C_1 \wedge C_2 \wedge ...C_n$ where each clause consists of three literals $C_i = x_{i1} \vee x_{i2} \vee x_{i3}, \forall i = 1, 2, .., n$.

we construct a matrix A like this: first, we construct m rows where each of them corresponds to a literal and then n rows where each of them corresponds to a clause. Then we construct 2m columns where first m columns corresponds to the literals x_i and the next m columns corresponds to literals $\neg x_i$. For the first m rows, we set $A_{i,i}$ and $A_{i,i+m}$ to 1 and the rest of the entries to 0; for the next m+n rows, we set the entries in row m+j corresponding to the each literal in clause c_j to 1 and the rest entries to 0. For example, given a formula, the construction of the matrix is shown in Fig. 1.

$C_1 = x_1 \vee x_2 \vee \neg x_3, C_2 = x_1 \vee \neg x_2 \vee x_3, ..., C_M = \neg x_1 \vee \neg x_2 \vee x_5$

	x_1	x_2	x_3	...	x_N	$\neg x_1$	$\neg x_2$	$\neg x_3$	...	$\neg x_N$
x_1	1	0		...	0	1	0	0	...	0
x_2	0	1		...	0	0	1	...	...	0
...	...	...		...	...	...	...	...	...	...
x_N	0	0		...	1	0	0	...	...	1
C_1	1	1	0				0	1		
C_2	1	0	1				1	0		
...	...	...		...	...	...	...	...	...	...
C_M				1		1	1			

⇨ A

Figure 1: An example for SAT reduction

Now we show that given any true instance of one problem, it corresponds to the true instance of the other problem:

"3SAT $\Rightarrow$ Fairness Maximization": given a positive instance of 3SAT, we know that each literal x_i is either 1 or 0, which means the first m rows satisfy the requirement; for the remaining n rows, in order to guarantee the clause to be true, at least one of the literal should be true, which means at least one entry of 1 should be selected. In summary, the requirement is satisfied.

"Fairness Maximization $\Rightarrow$ 3SAT": given a positive instance of MMFM, consider the first m rows, either the literal x_i or $\neg x_i$ is 1, which coincides with the satisfiability of any clause; For the remaining n rows, at least one entry with value 1 is selected, as the construction shows, the corresponding clause is true. Since each row corresponds to a clause, all the clauses are true, which makes the whole formula to be true. □

3.3 Problem Formulation

The scalar concept of "optimality" does not apply directly in the multi-objective setting. Here the notion of Pareto optimality is used to depict the optimality. Consider a solution S of multiple objective optimization problem, where each objective function is denoted as $f_i, i = 1, 2, ..., p$, therefore the values of the objective functions for solution S correspond to a vector: $(S_1, S_2, ..., S_p)$. A solution S' dominates solution S if $\forall i \in [1, p]$, $S_i \leq S'_i$ and $\exists j \in [1, p]$, $S_j < S'_j$.

Definition 3.5. **Pareto Optimal and Pareto Front:** A solution S is Pareto Optimal if and only if there exists no other solution S' that dominates S. The set of all Pareto optimal solutions is referred to as the Pareto Front.

Scarlarization is a typical method for solving a Pareto-Efficient solution for multiple objective solution. The scarlarization scheme is to assign weights to each objective and use the weighted sum of different objective functions as a single objective for proximity:

$$\lambda \cdot SW(g, I) + (1 - \lambda) \cdot F(g, I) \tag{1}$$

It is easy to see that when $0 < \lambda < 1$, the optimal solution to the scalarized single objective optimization problem is Pareto Optimal. Given that optimizing the single objective of

Algorithm 1 Greedy Algorithm for Fairness-Aware Recommendation to Groups

Input: A given group: G, A set of Items: I, Length of Recommendation List: K;

```
Initialize L = ∅;
while Length(L) < K: do
    Select an item j ∈ I \ L that maximizes λ · SW(g, L ∪ j) + (1 − λ) · F(g, L ∪ j);
    Add item j to the list L: L = L ∪ j;
end while
Output the recommendation list L;
```

fairness is NP-Hard, we look for efficient heuristic algorithms for the fairness-aware group recommendation problem.

4 OPTIMIZATION FRAMEWORK WITH PARETO EFFICIENCY

In this section, we formally introduce the optimization framework for fairness-aware group recommendation problem. Two ways are presented to achieve the Pareto-Efficient solutions, including scalarization with greedy algorithms and scalarization with integer programming.

Given the scalarized single objective optimization problem, we still face a NP-Hard problem as proved in last section. We propose a greedy algorithm for the scalarized problem which applies to all the proposed semantics of user utilities and fairness. The basic idea of greedy algorithm is to select an item that achieves the highest fairness when it is added to the current recommendation list. The detailed algorithm is illustrated in Alg. 1.

An important advantage of Greedy algorithm is the computational efficiency. Notice that solving an optimization problem can be time-consuming, it is even more complex to solve a non-convex optimization problem. However the Greedy algorithm gradually selects one item in each iteration, the algorithm runs K iterations to generate a final recommendation, which is more time-efficient.

Notice that some semantics for fairness are relatively simple in form, we can adopt integer programming techniques to solve the problem. We use $X_j \in \{0, 1\}, \forall j \in I$ to denote whether item $j \in I$ is recommended to the group. The recommendation process can be formulated into integer programming problems according to the individual utility and fairness semantics. We list the integer variable representation of the individual utilities:

- Average: $U(u, I) = \sum_{j \in I} rel(u, j) X_j$;
- Proportionality: $U(u, I) = \frac{\sum_{j \in I} rel(u,j) X_j}{\sum_{j \in I(u,k)} rel(u,j)}$.

Given these representations for individual utility functions, we can reformulate the problem of group recommendation into an integer program:

$$\begin{aligned} \max.\ & \lambda \cdot SW(g, I) + (1 - \lambda) \cdot F(g, I) \\ s.t.\ & \sum_i X_i = K \\ & X_i \in \{0, 1\} \end{aligned} \tag{2}$$

Algorithm 2 Integer Programming Based Algorithm for Fairness Aware Recommendation to Groups

Input: A given group: G, A set of Items: I, Length of Recommendation List: K;

1. Formulate the Recommendation problem as an integer programming following Eqn. 4;
2. Solve the Relaxed convex programming with Boundary Constraints, denote the solution as $X_j, \forall j \in I$;
3. Select K items I_m^K with greatest values of X_j;
4. Output the recommendation list $L = \{X_j \in I_m^K\}$;

For LM fairness semantic, the integer programming can be formulated as:

$$\begin{aligned} \max.\ & \lambda \cdot SW(g, I) + (1 - \lambda) \cdot \frac{1}{K} U_{min} \\ s.t.\ & \sum_i X_i = K \\ & U(u, I) \geq U_{min}, \forall u \in g \\ & X_i \in \{0, 1\} \end{aligned} \tag{3}$$

For Min-Max fairness semantic, the integer programming can be formulated as:

$$\begin{aligned} \max.\ & \lambda \cdot SW(g, I) + (1 - \lambda) \cdot \frac{U_{min}}{U_{max}} \\ s.t.\ & \sum_i X_i = K \\ & U(u, I) \geq U_{min}, \forall u \in g \\ & U(u, I) \leq U_{max}, \forall u \in g \\ & X_i \in \{0, 1\} \end{aligned} \tag{4}$$

For the variance and Jain's fairness semantics, the objective function can be directly reformulated by rewriting $F(u, I)$ with $U(u, I)$. The integer program is usually NP-Hard and very difficult to solve with an optimal solution. We can relax the binary constraint of integer variables to fractional variables with boundary constraints ($0 \leq X_i \leq 1$) and solve the relaxed program with a fractional solution. Then we round them into integers. Some semantics like Least Misery and Variance fairness ($\min_{u \in g} U(u, I)$ and $Var(\{U(u, I), \forall u \in g\})$) are relatively simple. Thus the objective function becomes convex and the relaxed program can be solved with an optimal solution. The algorithm based on integer programming is presented in Alg.2.

5 EXPERIMENT

We conduct extensive experiments on real-world datasets to evaluate our algorithms in terms of typical recommendation accuracy metrics. Moreover, we also exploit the impact of trade-offs on the item recommendation accuracy.

5.1 Experiment Settings

We choose the real-world datasets Movielens and MoviePilot for experiments. The Movielens dataset contains the interactions between individual users and items but has no groups, therefore we construct groups from the individual users with typical methods following previous work [2], [8]

Table 1: Statistics of the datasets.

Dataset	#Users	#Items	#Interactions	#Groups
Movielens	6,040	3,907	1,000,067	\
MoviePilot	171,670	23,974	4,391,822	290

and [21]. The users in MoviePilot dataset may share movie rental accounts with each other and those sharing same account form a group. There are a considerable number of users in MoviePilot dataset, only 602 users share accounts with others. Most of the groups contain only two users and the maximum size of the groups is four. The details of the datasets are listed in Table. 1.

In our experiments, we use five-fold cross-validation for Movielens where four folds are used for training and the remaining fold is used as testing set. The dataset of MoviePilot comes from the challenge of CAMRA and We thus use the training set to learn the model and rank all the non-rated items for each user and compare to the evaluation set for both the recommendation tasks. For datasets that contain only individual interactions with items, we generate three kinds of groups to conduct the experiments, including random user groups (RG), similar user groups (SG) and diverse user groups (DG). In order to generate these groups based on user similarities, we first compute the similarities of users based on their historical interactions with the items; then we randomly select G users for G groups and greedily select the most (least) similar user to each SG (DG) group from the remaining users. For each group, we set the group size as eight and 100 groups are generated for each dataset. We also alter the group size to evaluate its impacts.

We adopt the cosine similarity between the ratings given by two users to compute the similar and diverse user groups. For the missing individual preferences on items, we adopt state-of-the-art approach BPR [24] to make predictions on the Top-K favourite items of each user [1]. For the Top-K group recommendation task, we first compute the Top K most relevant items $I(u, K)$ for each individual user u and set their relevances with two semantics:

- Binary Semantic: $rel(u, i) = 1, \forall u, i \in I(u, k)$ and $rel(u, i) = 0, \forall u, i \notin I(u, k)$;
- Borda Semantic: $rel(u, i) = |\{j|rank(u, j) < rank(u, i), \forall j \in I(u, k)\}|$, $\forall u, i \in I(u, k)$, where $rank(u, i)$ is the ranking of item i in the Top-K personalized recommendation list for user u.

We also conduct experiments to show the performances of item recommendation to groups, some important metrics are adopted:

$$Rec@K = \frac{\sum_{i=1}^{K} rel_i}{\min(K, |y_u^{test}|)}; Prec@K = \frac{\sum_{i=1}^{K} rel_i}{K};$$

$$DCG@K = \sum_{i=1}^{K} \frac{2^{rel_i} - 1}{log_2(i+1)}; NDCG@K = \frac{DCG@K}{IDCG@K}$$

Meanwhile, we also adopt F score to evaluate the recommendation performances, which is a coordination of both precision and recall:

$$F@K = \frac{2Rec@K \cdot Prec@K}{Rec@K + Prec@K}$$

$rel_i = 1/0$ indicates whether the item at rank i in the Top-K recommendation list is in the test set. y_u^{test} denotes the items rated by user u in the testing set. The notion **IDCG** means the maximum possible **DCG** through ideal ranking.

We compare the performances of our algorithm with some existing state-of-art baselines, including:

- LM Ranking Algorithm [2]: this algorithm ranks the items based on the Least Misery relevances and recommend the Top-K items;
- Ave Ranking Algorithm [2]: this algorithm ranks the items based on the Average relevances and recommend the Top-K items;
- SPGreedy Algorithm [21]: this algorithm proposes a fairness metric called proportionality and greedily selects items to maximize the fairness;
- EFGreedy Algorithm [21]: this algorithm proposes a fairness metric called envy-freeness and greedily selects items to maximize the fairness;
- Greedy-LM (proposed in this paper): this algorithm is our proposed greedy algorithm for Least Misery fairness-aware group recommendation;
- Greedy-Var (proposed in this paper): this algorithm is our proposed greedy algorithm for Variance fairness-aware group recommendation.

Table 4: Performances Comparisons on MoviePilot with Borda Relevance, $K = 10$

Metrics	LM Ranking	Ave Ranking	SPGreedy	EFGreedy	Greedy-LM	Greedy-Var
Prec@K	0.0385	0.0467	0.0003	0.0013	0.0484	**0.0488**
Rec@K	0.0762	0.0910	0.0003	0.0020	0.0942	**0.0945**
F@K	0.0512	0.0617	0.0003	0.0016	0.0639	**0.0644**
NDCG@K	0.2376	0.2450	0.0008	0.0077	**0.2507**	0.2502

Table 5: Performances Comparisons on MoviePilot with Borda Relevance, $K = 20$

Metrics	LM Ranking	Ave Ranking	SPGreedy	EFGreedy	Greedy-LM	Greedy-Var
Prec@K	0.0423	0.0456	0.0008	0.0011	0.0464	**0.0465**
Rec@K	0.1483	0.1603	0.0015	0.0046	0.1636	**0.1639**
F@K	0.0658	0.0710	0.0010	0.0018	0.0723	**0.0724**
NDCG@K	0.2825	0.2901	0.0025	0.0080	0.2910	**0.2915**

5.2 Performance on Recommendation

We further investigate the performance of our algorithm on item recommendation tasks. We first utilize a state-of-art collaborative ranking method to generate the user's preference on unrecommended items as an ordering list. Then we adopt Proportionality as individual utility and different fairness metrics for recommendation. When making recommendations to the groups, an item may have been rated by one user while the others have not. We set the relevances of items to zeros for users who have rated them.

[1] We adopt the implementation of BPR from Librec [11]

Table 2: Performances of GreedyAlg-Var under different λ on MovieLens with Borda Relevance (Rand Group, Sim Group and Div Group), $|G| = 8$, $K = 10$

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
λ, SG	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
F@K	0.0103	0.0519	0.0631	0.0759	0.0772	0.0792	0.0799	0.0816	0.0821	0.0841	0.0819
NDCG@K	0.0335	0.1577	0.1691	0.1906	0.2003	0.2031	0.2072	0.2091	0.2079	0.2108	0.2079
λ, DG	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
F@K	0.0292	0.0656	0.0714	0.0848	0.0925	0.0950	0.0959	0.0978	0.0990	0.0983	0.0960
NDCG@K	0.0767	0.1715	0.1839	0.2076	0.2185	0.2246	0.2218	0.2237	0.2251	0.2248	0.2228

Table 3: Performances of GreedyAlg-Var under different λ on MovieLens with Binary Relevance (Rand Group, Sim Group and Div Group), $|G| = 8$, $K = 10$

λ, RG	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
F@K	0.0193	0.0452	0.0477	0.0596	0.0851	0.0985	0.1015	0.1032	0.1047	0.1059	0.1017
NDCG@K	0.0584	0.1505	0.1547	0.1722	0.2076	0.2322	0.2406	0.2377	0.2389	0.2419	0.2357
λ, SG	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
F@K	0.0088	0.0294	0.0328	0.0479	0.0678	0.0725	0.0739	0.0761	0.0785	0.0786	0.0757
NDCG@K	0.0317	0.1102	0.1168	0.1442	0.1720	0.1803	0.1836	0.1872	0.1923	0.1931	0.1834
λ, DG	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
F@K	0.0202	0.0414	0.0433	0.0551	0.0753	0.0879	0.0893	0.0884	0.0887	0.0895	0.0889
NDCG@K	0.0574	0.1386	0.1441	0.1623	0.1904	0.2063	0.2132	0.2094	0.2092	0.2052	0.2051

We select the values of λ uniformly from the interval $[0, 1]$ and generate recommendation lists for different scenarios. A 10% sample of the dataset is used as validation set to choose the parameter λ. We compare the performances of comparative algorithms with different semantics of relevances on three types of groups and the results are presented in Fig. 2 and Fig. 3.

The experimental results show that considering fairness in item recommendation leads to better accuracy, regardless of which type of groups is considered. It is an interesting finding since fairness is not directly related to recommendation accuracy. The possible reason is that considering fairness can make up for the imbalance between users' satisfactions and users who are less satisfied can get higher utilities so that the overall accuracy is improved. Meanwhile, we find that the improvements of recommendation accuracy are not the same on all types of groups.

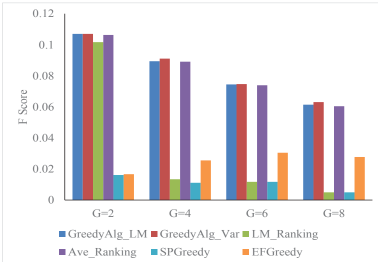
We also compare the performances under different choices of λ, the results are presented in Table. 3. The variance fairness maximization tries to minimize the difference between the utilities of users and considering the fairness at a proper level is important for recommendation accuracy. The empirical performances show that setting λ to 0.8 to 0.9 leads to the best results. Since maximizing fairness alone may harm the satisfaction of users (when recommending commonly disliked items to the group, the fairness is high but the overall social welfare is low), this finding is reasonable.

We also conduct experiments on the MoviePilot dataset where users can share a same movie rental account and those users form a movie rental group. Notice that the groups in this dataset usually contain only two users, therefore the setting is closer to personalized recommendation than the setting of Movielens. Comparing the performances of group recommendation with binary and borda semantics for relevance, we find that using the borda semantic preserves the comparative differences between items thus leads to better performances, especially when the group size is small. Due to the better performances on borda semantic on small groups, we present the results of MoviePilot with borda semantic with $K = 10$ and $K = 20$ in Table. 4 and Table. 5.

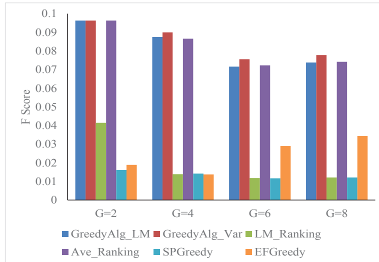
The results from the table indicate that our algorithms still achieve superior performances. Since the groups in MoviePilot are real, the results show that the group decision process does involve the coordination of individual users and considering fairness in the process can improve the recommendation accuracy. As the groups in MoviePilot are small and the improvements of our algorithm are more significant when groups are large in size, our algorithm can be more useful in other real-life groups whose sizes are larger.

6 CONCLUSION

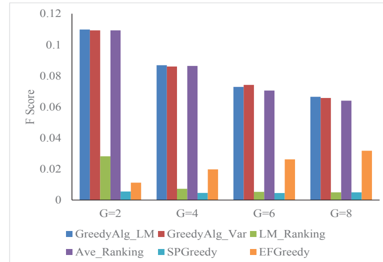
In this paper, we investigate how fairness can be modeled in group recommendation and its impact on the quality of recommendation. More specifically, we propose various semantics for user individual utility and further model fairness as a proximity of how balanced the utilities of users are when group recommendations are given. Based on the proposed utilities and fairness, we consider the fairness-aware group recommendation problem and prove its NP-Hardness. In order to optimize both user utility and fairness in group recommendation, we propose a general optimization framework based on Pareto Efficiency. We conduct extensive experiments on real-world datasets to evaluate the performance of



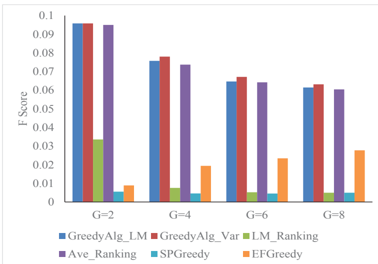
(a) Borda Relevance, Rand Groups



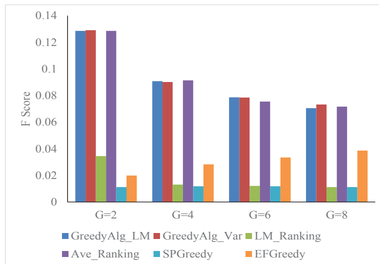
(b) Binary Relevance, Rand Groups



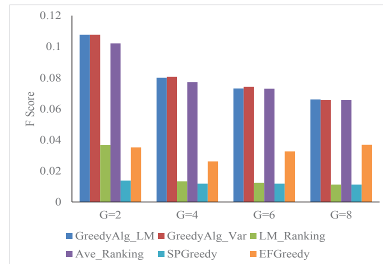
(c) Borda Relevance, Sim Groups



(d) Binary Relevance, Sim Groups

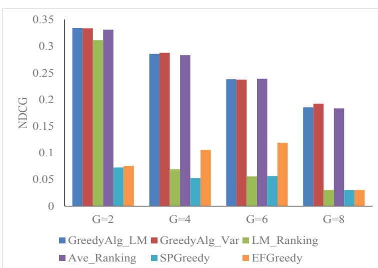


(e) Borda Relevance, Div Groups

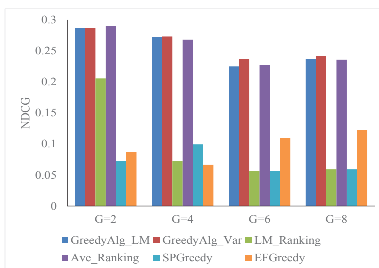


(f) Binary Relevance, Div Groups

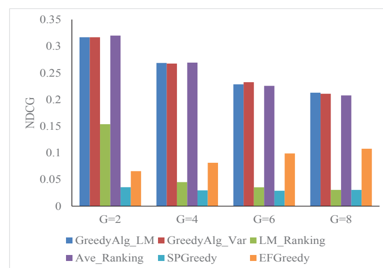
Figure 2: F@10 with different group sizes on MovieLens



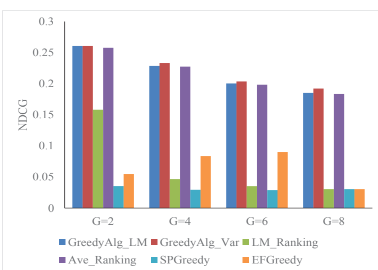
(a) Borda Relevance, Rand Groups



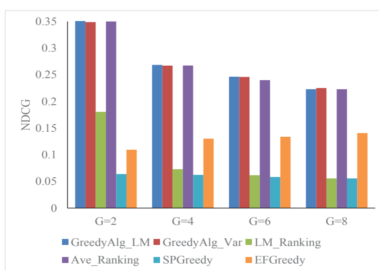
(b) Binary Relevance, Rand Groups



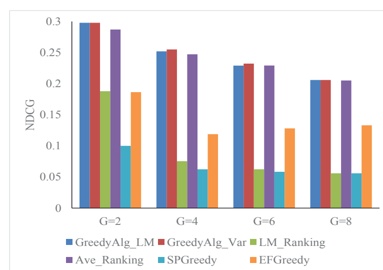
(c) Borda Relevance, Sim Groups



(d) Binary Relevance, Sim Groups



(e) Borda Relevance, Div Groups



(f) Binary Relevance, Div Groups

Figure 3: NDCG@10 with different group sizes on MovieLens

group recommendation in terms of accuracy. The results indicate that considering fairness can improve the quality of group recommendation.

7 ACKNOWLEDGEMENT

This work is supported by Natural Science Foundation of China (Grant No. 61532011, 61672311) and National Key Basic Research Program (2015CB358700). Part of the work is sponsored by the Center for Intelligent Information Retrieval. Meanwhile it is supported in part by the Hong Kong Scholars Program.

REFERENCES

[1] S. Amer-Yahia, S. B. Roy, A. Chawlat, G. Das, and C. Yu. Group recommendation: Semantics and efficiency. *Proc. VLDB Endow.*, 2(1), Aug. 2009.

[2] L. Baltrunas, T. Makcinskas, and F. Ricci. Group recommendations with rank aggregation and collaborative filtering. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 119–126. ACM, 2010.

[3] F. Brandt, V. Conitzer, and U. Endriss. Computational social choice. *Multiagent systems*, pages 213–283, 2012.

[4] L. A. Carvalho and H. T. Macedo. Generation of coalition structures to provide proper groups' formation in group recommender systems. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 945–950. ACM, 2013.

[5] L. A. M. C. Carvalho and H. T. Macedo. Users' satisfaction in recommendation systems for groups: an approach based on noncooperative games. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 951–958. ACM, 2013.

[6] D. Cosley, J. A. Konstan, and J. Riedl. Polylens: A recommender system for groups of user. In *In ECSCW*, 2001.

[7] A. Crossen, J. Budzik, and K. J. Hammond. Flytrap: Intelligent group music recommendation. In *IUI '02*, New York, NY, USA, 2002. ACM.

[8] T. De Pessemier, S. Dooms, and L. Martens. Design and evaluation of a group recommender system. In *Proceedings of the sixth ACM conference on Recommender systems*, pages 225–228. ACM, 2012.

[9] K. Deb, K. Sindhya, and J. Hakanen. Multi-objective optimization. In *Decision Sciences: Theory and Practice*, pages 145–184. CRC Press, 2016.

[10] M. Gartrell, X. Xing, Q. Lv, A. Beach, R. Han, S. Mishra, and K. Seada. Enhancing group recommendation by incorporating social relationship interactions. In *Proceedings of the 16th ACM international conference on Supporting group work*, pages 97–106. ACM, 2010.

[11] G. Guo, J. Zhang, Z. Sun, and N. Yorke-Smith. Librec: A java library for recommender systems. In *UMAP Workshops*, 2015.

[12] F. Guzzi, F. Ricci, and R. Burke. Interactive multi-party critiquing for group recommendation. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 265–268. ACM, 2011.

[13] L. Hu, J. Cao, G. Xu, L. Cao, Z. Gu, and W. Cao. Deep modeling of group preferences for group-based recommendation. In *AAAI*, pages 1861–1867, 2014.

[14] R. Jain, D.-M. Chiu, and W. R. Hawe. *A quantitative measure of fairness and discrimination for resource allocation in shared computer system*, volume 38. Eastern Research Laboratory, Digital Equipment Corporation Hudson, MA, 1984.

[15] T. Jambor and J. Wang. Optimizing multiple objectives in collaborative filtering. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 55–62. ACM, 2010.

[16] M. Kompan and M. Bieliková. Voting based group recommendation: How users vote. In *HT (Doctoral Consortium/Late-breaking Results/Workshops)*, 2014.

[17] X. Liu, Y. Tian, M. Ye, and W.-C. Lee. Exploring personal impact for group recommendation. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 674–683. ACM, 2012.

[18] L. Naamani Dery, M. Kalech, L. Rokach, and B. Shapira. Iterative voting under uncertainty for group recommender systems. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 265–268. ACM, 2010.

[19] E. Ntoutsi, K. Stefanidis, K. Nørvåg, and H.-P. Kriegel. Fast group recommendations by applying user clustering. In *ER'12*, Berlin, Heidelberg, 2012. Springer-Verlag.

[20] W. Pan and L. Chen. Gbpr: Group preference based bayesian personalized ranking for one-class collaborative filtering. In *IJCAI*, volume 13, pages 2691–2697, 2013.

[21] S. Qi, N. Mamoulis, E. Pitoura, and P. Tsaparas. Recommending packages to groups. In *Data Mining (ICDM), 2016 IEEE 16th International Conference on*, pages 449–458. IEEE, 2016.

[22] E. Quintarelli, E. Rabosio, and L. Tanca. Recommending new items to ephemeral groups using contextual user influence. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 285–292. ACM, 2016.

[23] J. A. Recio-Garcia, G. Jimenez-Diaz, A. A. Sanchez-Ruiz, and B. Diaz-Agudo. Personality aware recommendations to groups. In *Proceedings of the third ACM conference on Recommender systems*, pages 325–328. ACM, 2009.

[24] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*, pages 452–461. AUAI Press, 2009.

[25] M. T. Ribeiro, A. Lacerda, A. Veloso, and N. Ziviani. Pareto-efficient hybridization for multi-objective recommender systems. In *Proceedings of the sixth ACM conference on Recommender systems*, pages 19–26. ACM, 2012.

[26] F. Ricci, L. Rokach, and B. Shapira. *Introduction to recommender systems handbook*. Springer, 2011.

[27] M. Rodriguez, C. Posse, and E. Zhang. Multiple objective optimization in recommender systems. In *Proceedings of the sixth ACM conference on Recommender systems*, pages 11–18. ACM, 2012.

[28] I. Ronen, I. Guy, E. Kravi, and M. Barnea. Recommending social media content to community owners. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 243–252. ACM, 2014.

[29] S. B. Roy, S. Thirumuruganathan, S. Amer-Yahia, G. Das, and C. Yu. Exploiting group recommendation functions for flexible preferences. In *Data Engineering (ICDE), 2014 IEEE 30th International Conference on*, pages 412–423. IEEE, 2014.

[30] A. Said, S. Berkovsky, and E. W. De Luca. Group recommendation in context. In *Proceedings of the 2nd Challenge on Context-Aware Movie Recommendation*, pages 2–4. ACM, 2011.

[31] A. Salehi-Abari and C. Boutilier. Preference-oriented social networks: Group recommendation and inference. In *RecSys '15*, New York, NY, USA, 2015. ACM.

[32] A. Salehi-Abari and C. Boutilier. Preference-oriented social networks: Group recommendation and inference. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 35–42. ACM, 2015.

[33] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on World Wide Web*, WWW '01, pages 285–295, New York, NY, USA, 2001. ACM.

[34] S. Seko, T. Yagi, M. Motegi, and S. Muto. Group recommendation using feature space representing behavioral tendency and power balance among members. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 101–108. ACM, 2011.

[35] A. Sen. Social choice theory. *Handbook of mathematical economics*, 3:1073–1181, 1986.

[36] P. Skowron, P. Faliszewski, and J. Lang. Finding a collective set of items: From proportional multirepresentation to group recommendation. *Artificial Intelligence*, 241:191–216, 2016.

[37] X. Wang, R. Donaldson, C. Nell, P. Gorniak, M. Ester, and J. Bu. Recommending groups to users using user-group engagement and time-dependent matrix factorization. In *AAAI*, pages 1331–1337. AAAI Press, 2016.

[38] M. Ye, X. Liu, and W.-C. Lee. Exploring social influence for recommendation: a generative model approach. In *SIGIR*, pages 671–680. ACM, 2012.

[39] Q. Yuan, G. Cong, and C.-Y. Lin. Com: a generative model for group recommendation. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 163–172. ACM, 2014.

[40] W. Zhang, J. Wang, and W. Feng. Combining latent factor model with location features for event-based group recommendation. In *KDD*, pages 910–918. ACM, 2013.

[41] E. Zitzler and L. Thiele. Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach. *IEEE transactions on Evolutionary Computation*, 3(4):257–271, 1999.