A Group Recommendation Model Using Diversification Techniques

Amanda Chagas de Oliveira
Universidade Federal da Bahia
oliveira.amanda@ufba.br

Frederico Araújo Durão
Universidade Federal da Bahia
fdurao@ufba.br

Abstract

In daily life groups are formed naturally, such as watching a movie with friends, or going out for dinner. In all these scenarios, using Recommendation Systems can be helpful by suggesting pieces of information (e.g. movies or restaurants) that satisfies all rather than a single member in the group. To do so, it is crucial to aggregate individual preferences of the group members aiming at satisfying all. Although there are consensus techniques to create the group profile, the recommendations still may be repetitive and overspecialized. This drawback sets precedent for adopting diversification techniques to group recommendations. In this paper, we propose a group recommendation model using diversification techniques that exploits different aggregation techniques over group preferences matrix. The experiments evaluate accuracy and diversity goals for the group recommendations. Results from the experiments point out that our approach achieved 1.8% of diversity increase and 3.8% of precision improvement over compared methods.

1. Introduction

Recommender Systems (RS) can be defined as an automated tool for the process of seeking what is relevant for the user [1]. Most of RS are designed to satisfy individuals instead of a group of users. However, there are scenarios where the recommendations should be made for a group. A Group Recommender System (GRS) has to consider preferences of each group member in the recommended items, and therefore, help the members choose what items are best for them.

Group Recommender Systems (GRS) have the role of finding what is relevant to the group rather than the individuals. Considering that we live in communities, engaging in group activities is a natural behavior. However, a simple task such as choosing a playlist for a road trip may be challenging. In a group scenario, all individual preferences must be considered when recommendations are processed. At least, the recommendation problem is multiplied by the size of the group. The preference of each member must be considered when building the group profile without prioritization. Hence, consensus techniques must be employed in the way of finding which items are relevant not only for a single member but for the entire group as a whole. GRS have been studied in the literature from various aspects. [1] investigate strategies for aggregating individual preferences, [2] focus on efficient ways for group recommendations, and [3] leverage different approaches when recommending items to groups.

Group recommendations rely on a group profile, which is based on an aggregation of individuals' preferences. If such a profile is not updated or revised, recommendations tend to fall into regions the group members are familiar with, without experiencing any sort of diversity. This problem, known as overspecialization, may impact GRSs negatively since recommendations may become repetitive, thus unattractive. Hence, it is important that a GRS exploit different approaches to diversify recommendations with the goal of improving the overall group satisfaction. In industry, the music streaming service Spotify `www.spotify.com` has added a playlist in its recommendation called "Family Mix", which combines songs of members subscribed in the same Family plan. No information about the diversity algorithm is exposed though. In the literature, diversity in RS has been widely investigated. [4] was the first work to formalize the idea of diversity, [5] investigate diversity when evaluating recommender systems, and [6] discuss the impact of diversity algorithms in group recommendations.

This work focuses on tackling the overspecialization problem using diversity for group recommendations. This approach proposes the development of a group recommendation model that applies a diversification algorithm aiming at optimizing consensus among members besides increasing plurality in the

URI: https://hdl.handle.net/10125/70940
978-0-9981331-4-0

recommendations. The major goal therefore is to increase the satisfaction of the group members by decreasing the impact of overspecialization.

This paper is structured as follows. Section 2 presents related work. Section 3 provides important background on the area, Section 4 depicts the overall approach, Section 5 presents the experimental evaluation and Section 6 concludes the paper and sets forth the future works.

2. Related Work

A Group Recommendation System (GRS) suggests items to a group of people engaged in a group activity. In the literature, group recommendation has been widely studied from many perspectives such as aggregation techniques [1, 5], group consensus [3] and polarization [7]. Nevertheless, the adoption of diversity algorithms in GRS is still an open field [6].

As to GRS, [1] provide a comprehensive overview and meaningful definitions about several aggregation techniques, including *Least Misery*, *Most Pleasure* and *Average Without Misery*. [2] propose semantics that account item relevance and group disagreements, but also use three group formation approaches to binding users together: first similar users, then dissimilar users, and finally groups formed randomly. In our work, we create groups randomly as detailed in Section 4.2. Also, [7] analyze the existence of group polarization effects in GRS and propose a way to counteract these effects. The authors perform a user trial focusing on risk analysis. Despite the importance of group polarization in group decisions, this problem is not addressed in our work, as we focus on modeling groups and diversification of recommendations. [8] investigate a hybrid GRS that combines content-based and collaborative strategies. Similarly to our approach, the authors try to predict ratings for creating a group profile, and then generate recommendations using a content based approach. [9] propose a method that predicts user ratings and uses a graph aggregation strategy to bind those preferences of the group members. Unlike our work, they discards diversity as a mean to improve satisfaction among the group members.

Diversity in Recommender System is discussed by [5], which outlines important issues beyond-accuracy in RS. In particular, they discuss the benefits of diversification algorithms and present several approaches about re-ranking lists. Likewise, we also opted for this approach however from a group perspective. Diversity is a large concept, the authors in [10] propose a method for diversifying Explanation Style (ES) on recommendation and developed a framework called DualDiv that can increase the diversity of the items and the ES with no big impact on accuracy. In addition, [11] proposes a Community-Aware Diversification (CAD) approach that is based in an intent-aware diversification method called Subprofile-Aware Diversification (SPAD) [12]. The authors of [11] conclude their proposal is equivalent or even better than compared baseline methods. Finally, authors of [6] also discuss the problem of diversity in GRS, nevertheless it is unclear how their preference matrix is constructed.

3. Background

3.1. Aggregation Techniques

In a GRS, aggregation techniques are consensus functions capable of combining multiple preferences in a single one, i.e. the group profile. The approaches of aggregation are [3]:

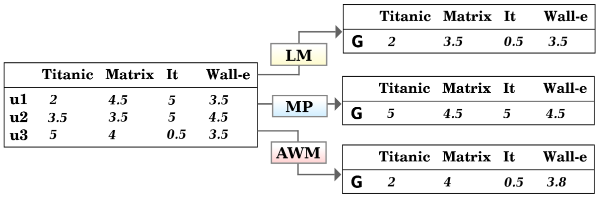
- **Merging of Recommendations Made for Individuals.** In this approach, a list of recommendations for each member is generated individually, and then those lists are merged into a single one.

- **Aggregation of Ratings for Individuals.** This is the most common approach in the literature, it assumes that for each candidate item, the system can predict the rating of each member of the group and aggregate those values applying aggregation techniques.

- **Construction of Group Preference Models.** This is the second most common approach, it is not based on prediction but rather uses individual preferences of the members to construct a group model as a whole.

The proposed approach aggregates individual ratings to reach a group profile. Inspired by [1], this approach relies on three aggregation techniques: Least Misery (LM), Most Pleasure (MP), and Average Without Misery (AWM). Figure 1 depicts these techniques.

- **Least Misery (LM).** Given the members' ratings over each candidate item, this technique assumes the least individual rating as the group rating for that candidate item. The main idea is to avoid misery for group members. For example, if the items can be rated from 0.5 to 5, and the LM indicates an item that scored 4, it means that the member who gave the least score still is pleased.



	Titanic	Matrix	It	Wall-e
u1	2	4.5	5	3.5
u2	3.5	3.5	5	4.5
u3	5	4	0.5	3.5

	Titanic	Matrix	It	Wall-e
G	2	3.5	0.5	3.5

	Titanic	Matrix	It	Wall-e
G	5	4.5	5	4.5

	Titanic	Matrix	It	Wall-e
G	2	4	0.5	3.8

Figure 1: Ensemble aggregation techniques individual's preferences.

- **Most Pleasure (MP).** The MP technique is the opposite of LM. Given the members' ratings over each candidate item, this technique assumes the highest individual rating as the group rating for that candidate item. This technique tries to keep an item that is highly relevant to a member of the group. However, it is possible to recommend an item that is badly scored by another member, being unpleasant to the group as a whole.

- **Average Without Misery (AWM).** This technique can be described as a combination of two other techniques: Average and Least Misery. The Average technique consists in calculating the mean of individuals ratings over each candidate item and assume the mean as the group rating for that candidate item. In this work, the AWM was implemented as such: for each candidate item , if there is an individual rating equal or under the threshold, LM defines the group rating for that candidate item. However, if the individual ratings are over the threshold, then Average defines the group rating for that candidate item. Considering the dataset used in this work, in which ratings vary from 0.5 to 5, we define the threshold as 2 based on the Likert Scale [13]. As 2 in this scenario is the value that already indicates displeasure. Figure 1 illustrates this scenario.

The Average and Least Misery techniques are present in most works studied in the literature. Thus, we implemented the AWM as the combination of both. The LM technique is implemented aiming for a better evaluation with other systems. The MP technique was implemented to serve as contrast for the LM.

3.2. Diversification Algorithms

Diversification can be defined on a list of items: a_1, a_2,..., a_k as a factor indicating how different pair items are on this list [4].

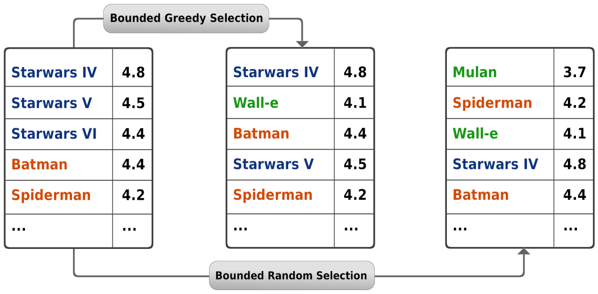
$$Similarity(x, y) = \frac{\sum_{i=1}^{n} w_i \cdot sim_i(x_i, y_i)}{\sum_{i=1}^{n} w_i} \qquad (1)$$

The similarity between a pair of items is defined in Equation 1, where n is the item features, w is the weight of the feature, and $sim(x, y)$ is the comparison of feature i from items x and y. The similarity calculus used in this work is the Cosine Similarity and we use the **title** and **genres** as features from the items.

$$Diversity(R) = \frac{\sum_{x \in R} \sum_{y \in R/\{x\}} dist(x, y)}{|R| \cdot (|R| - 1)} \qquad (2)$$

The diversity score is calculated accordingly with the Equation 2 [4], where R is the list of items, x e y represent items of the list, and $dist(x, y)$ is the distance function between those items, which also can be expressed as $1 - Similarity(x, y)$.

In particular, this proposal evaluated two diversification approaches from the literature: Bounded Random Selection and Bounded Greedy Selection. Figure 2 shows the outputs from these algorithms given a list of recommendations. Additionally, we also evaluated the diversification approach from the baseline [6].



Starwars IV	4.8
Starwars V	4.5
Starwars VI	4.4
Batman	4.4
Spiderman	4.2
...	...

Starwars IV	4.8
Wall-e	4.1
Batman	4.4
Starwars V	4.5
Spiderman	4.2
...	...

Mulan	3.7
Spiderman	4.2
Wall-e	4.1
Starwars IV	4.8
Batman	4.4
...	...

Figure 2: Ensemble diversification algorithms over recommendations.

- **Bounded Random Selection.** On this algorithm, there is a list L with the preferred items of the user, a list of candidate items C and the final list with diversified recommendation R. For each item i_l in L, the algorithm searches for items in C similar to i_l and adds those items in a new list J, with a bounded length. Then, items are chosen randomly in J and added in R.

- **Bounded Greedy Selection.** This algorithm selects items in a greedy way, by picking the most diverse item on each turn and adding it on R. However, it is crucial to define a function for the picking. This greedy function needs to consider both, similarity and diversity of the items. The flow of this algorithm is similar to the Random, however, instead of choosing randomly items from J, the greedy function picks those items that maximize diversity still considering similarity.

- **Baseline Diversification Approach.** This approach aims at maximizing item satisfaction to group and minimizing the similarity between recommendations. This function returns a list of item I_G of k items that maximize the objective function [6]:

$$I_G = \arg\max_{I \in \mathcal{D}} g(I), \tag{3}$$

where D contains all items and $g(I)$ is defined as:

$$g(I) = w \sum_{d \in I} q(d) r(d) - \sum_{d,d' \in I} r(d) S(d, d') r(d'), \tag{4}$$

where w is a positive parameter that defines the trade-off between two terms, I is the list of recommended items returned. The first term of this function measures the weighted utility of items in I. The second term introduces and measures the similarity of items in I.

$$\begin{aligned} Diversifying(x, R) &= \alpha \cdot rel(x) + ... \\ ... + (1-\alpha) &\cdot \frac{1}{|R|} \sum_{y \in R} dist(x, y) \end{aligned} \tag{5}$$

In this work, the greedy function for weighting diversity and similarity is expressed in Equation 5 [5], where α is a variable to balance factors of the equation, and $rel(x)$ is the relevance function for item x, which can be expressed based on similarity. This equation is a representation of the greedy diversification studied in [14], [15], and [5].

4. A Group Recommendation Model Using Diversification Techniques

4.1. Notations and Proposal Flow

The proposed model recommends a movie $i \in I$ to a group G composed of n users $u \in U$. Figure 3 illustrates a group G composed by 5 members, u_1, u_2, u_3, u_4, and u_5, being recommended with the item i, that is the movie *Moulin Rouge*.

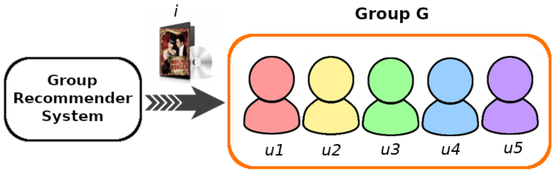


Figure 3: Illustration of a group G with n=5 users.

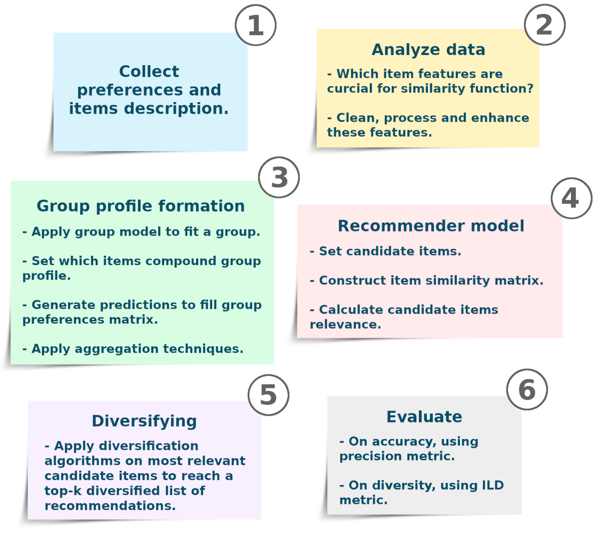

1
Collect preferences and items description.
2
Analyze data
- Which item features are curcial for similarity function?
- Clean, process and enhance these features.
3
Group profile formation
- Apply group model to fit a group.
- Set which items compound group profile.
- Generate predictions to fill group preferences matrix.
- Apply aggregation techniques.
4
Recommender model
- Set candidate items.
- Construct item similarity matrix.
- Calculate candidate items relevance.
5
Diversifying
- Apply diversification algorithms on most relevant candidate items to reach a top-k diversified list of recommendations.
6
Evaluate
- On accuracy, using precision metric.
- On diversity, using ILD metric.

Figure 4: Flow of proposed model defined in steps.

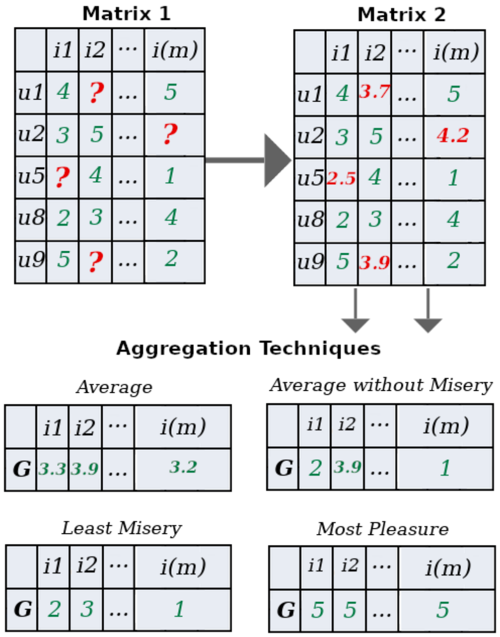
Figure 4 presents the recommendation flow. The first step is related to obtaining information from the dataset. On the next step, that information is cleaned and prepossessed aiming to reduce noise and improve similarity calculus between items. On Step 3, the group is formed accordingly with a group formation model. In a sequence, the aggregation technique is applied to create the group profile. Then, the recommendations are generated as illustrated in Step 4. Diversification occurs in Step 5. Finally, Step 6 is responsible for evaluating the group recommendations. In the following, the Steps 3, 4, and 5 are particularly discussed as follows.

4.2. The Group Model

The concept of a group can be defined as a system of recurrent social relations, or as a reunion of people who share some characteristic, some idea, or some common interest [16]. Therefore, a group can refer to some friends going out to watch a movie. In any case, it is possible to extract something that binds those people together. Thus, when recommending to groups, it is crucial to define clear rules for group formation. We divided the rules into two: group size and cohesion between members.

- **Group size.** The size of the group can vary drastically even in real life, from a couple having lunch to a crowd of thousands in a football stadium. With that in mind, we set the group size to 5 inspired also by recurrent use in the literature [8] and [6].
- **Cohesion between members.** This aspect focuses on the relationship between members. Most of the datasets used in RS focuses on modeling individuals' preferences rather than the relationship between those users. In this paper, the group members are chosen randomly.

Once the group is created, the next step is to define the group profile. In this phase, aggregation techniques are applied to individual preferences. Nevertheless, the group preferences matrix must be created, which consists of items that are rated by at least one group member. Figure 5 illustrates such a matrix.



Matrix 1

	i1	i2	···	i(m)
u1	4	?	...	5
u2	3	5	...	?
u5	?	4	...	1
u8	2	3	...	4
u9	5	?	...	2

Matrix 2

	i1	i2	···	i(m)
u1	4	3.7	...	5
u2	3	5	...	4.2
u5	2.5	4	...	1
u8	2	3	...	4
u9	5	3.9	...	2

Aggregation Techniques

Average

	i1	i2	···	i(m)
G	3.3	3.9	...	3.2

Average without Misery

	i1	i2	···	i(m)
G	2	3.9	...	1

Least Misery

	i1	i2	···	i(m)
G	2	3	...	1

Most Pleasure

	i1	i2	···	i(m)
G	5	5	...	5

Figure 5: Group profile generation flow using aggregation techniques.

Figure 5 presents the group with 5 members u_1, u_2, u_5, u_8, and u_9, and those respective ratings over profile items i_1 to $i_{(m)}$. The ratings range from 0.5 to 5 and the symbol *?* indicates no rating. In order to apply aggregation techniques at the matrix, we fill the vacant spaces (*?*) of the group preferences matrix with a prediction for u over i. To do that, we tested different prediction algorithms: neighborhood-based [17] i) *KnnWithMeans*, ii) *Knn*; and matrix factorization based iii) Singular Value Decomposition (SVD) [18]. We then selected the SVD as which performed best.

Figure 5 represents the dense group preferences matrix filled in Matrix 2. Then, the last step on the group profile creation is to apply the aggregation techniques. Figure 5 illustrates some aggregation techniques considered in our approach, such as, AWM, LM, and MP. Then we have a representative value $r_{G,i}$ for the group over the profile item i.

4.3. Recommendation Model

Algorithm 1: Algorithm for Group Recommendation.

```
input: $n$ as the number of groups;
       $size$, where $size = 5$;
       $Users$ as all dataset users;
       $sim$ as the similarity matrix of items.
while $n > 0$ do
    G = random(Users, size)
    pi, ci = splitDatasetItems(Items, G)
    Gp = generateGroupProfile(pi, G)
    bc = getMostSimilarItems(Gp, ci, sim)
    for $x \in bc$ do
        for $y \in Gp$ do
            $r = relevance(x, y)$
            $R = R + r$
        end
    end
    dr = diversifyRecommendations(R)
    n = n - 1
end
```

Generally, datasets in RS comprise data from individuals rather than groups of users. Moreover, even if we generate a group, there is no reference to what other groups have rated before in order to compare group recommendations. Therefore, it is impracticable to use the Collaborative Filtering (CF) approach on recommendation to groups. For this reason, we employed the Content Based (CB) approach to generate the recommendations to the groups, as it depends on the similarity between items. The CB approach [19] focuses on recommending items similar to previous items rated by the group. Thus, provided the group profile, we recommend unknown items that are similar to the better-rated items on the group profile.

Algorithm 1 illustrates how the recommendations are generated.

In Algorithm 1 pi and ci stands for profile items and candidate items respectively. Gp is the group profile and it is sorted by ratings, bc refers to best candidates and stores the most similar candidate items from profile items. The list R is for the recommendations generated to the group, and dr is the final list with the recommendations diversified. The recommendation model developed in this proposal follows this flow:

1. **Set candidate items.** It is important to split the items of the dataset between candidates and profiles. As mentioned in Section 4.2, any item that was rated by a group member is classified as a profile item. In other words, the rest of the items are unknown for the group as a whole, and those are the candidate items to be recommended. This step is defined in line 3 of the algorithm, where pi are the profile items and ci are the candidate items.

2. **Sort group profile on a higher rating.** Once the group profile is settled, we know which items the group enjoys the most. In this proposal, we sort the profile item as a decreasing $r_{G,i}$ in order to keep the preferred items at the beginning of the list. This step is defined in line 4 of the algorithm.

3. **Construct items similarity matrix.** The next step is to calculate the similarity between dataset items. We reach a similarity matrix where the lines and columns are the items, and the cell value (x, y) stands for the similarity score for items x and y. Particularly for the experiments we use the Cosine Similarity function [20] as it performs great with textual information. The experiment for weighing the cosine similarity was essentially based on the relevance of the returned items, i.e. which movies are more similar to the group profile. We performed several tests empirically until achieving the best relevance. Then we applied the following setting: 0.8 for title and 0.2 for genre. Algorithm 1 receives the similarity matrix as input sim.

4. **Setting relevance.** At this stage on the model, we already have the group score over profile items, and we know which candidate items are more similar to profile items bc. Therefore, it is crucial to quantify how relevant a candidate item is to a group. Thus, we elaborate the Equation 6 that combines similarity values with group preference to express the relevance of the item.

$$relevance(x, y) = \frac{a \cdot sim(x, y) + b \cdot \frac{r_{(G,y)}}{max(r)}}{a + b} \tag{6}$$

Equation 6 shows x as a candidate item from bc and y as a preferred item from Gp for group G. In the first factor of the equation, the similarity between items, $sim(x, y)$, is normalized, as well as the second factor in Equation 6 $r_{G,y}$. Then, the relevance of x is the result of the weighted mean of variables a and b over x and y similarity, and how preferred y was rated by G, respectively. The sum of variables a and b may never be less than 1. This step is defined in line 8 of the algorithm.

4.4. Diversifying Group Recommendations

After applying the group model and the recommendation model, we have a list of candidate items that are relevant for the group $recs$, however, how diverse are those items? In other words, how similar the top-k items are on this list? Considering k as 10, most items tend to have similar aspects between each other. Therefore, we diversify the top-k aiming to present as relevant and diverse items as possible to the group.

Most of the diversification algorithms are based on re-ranking items to improve diversity. In this model, we implement two of those algorithms: the *Greedy Re-ranking Algorithm* [5] and the *Bounded Random Selection* [4]. Those approaches of diversification are used in Algorithm 1 on line 12.

Figure 6 illustrates recommendations for group G, using traditional recommendation techniques versus diversified recommendations. Observing the traditional recommendation, up to the top-5 items, movies belong to the same genre, i.e. thriller, and a different genre is only observed in the seventh item. In contrast, some diversity is already observed in the top-3 list. It is important to outline that both lists are formed by the same items, being the second an output from a diversification process.

5. Experimental Evaluation

The experiments aim at evaluating the precision and diversity of the proposed model in comparison with state-of-the-art related works.

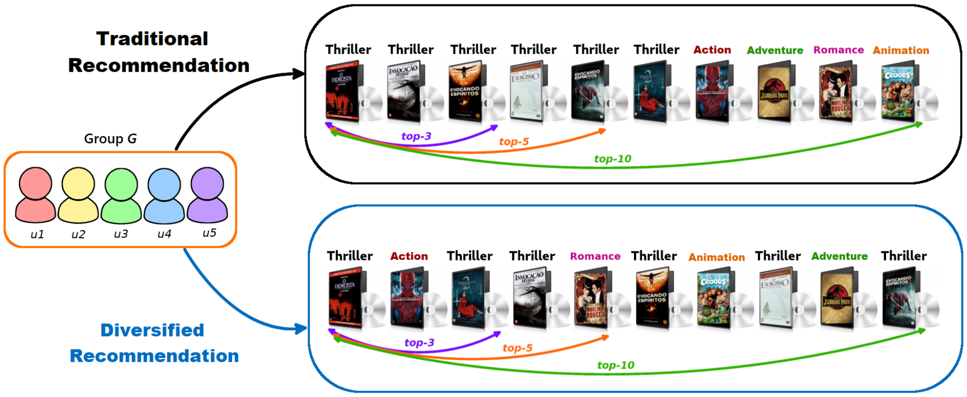


Figure 6: Impact of diversification on top-k.

5.1. Methodology

In order to assess the performance of our approach, we compared our model against the following existing approaches:

1. **Baseline (BL):** Diversifying Group Recommendation [6].
2. **Standard (SD):** Recommendation using only traditional recommendation techniques without diversification.
3. **Greedy (GD):** Diversifying using the Greedy algorithm.
4. **Random (RD):** Diversifying using the Bounded Random algorithm.

Besides comparing our approach against others, we also conducted an empirical evaluation in order to evaluate the quality of our group profile. Because our proposal depends on a dense matrix, we evaluated the best prediction method (see Section 4.2) before performing the aggregation techniques. The SVD method achieved the best predictions and was selected.

Provided that, we evaluated the recommendations in terms of precision and diversity. For that, we analyze the Precision@k and the Intra List Diversity (ILD) metrics. In this experiment, the aggregation techniques considered were Least Misery (LM), Most Pleasure (MP), and Average Without Misery (AWM). We tested the recommendations for five groups randomly created, each containing five users. Grouping people randomly is an approach used to avoid bias in group formation [6]. Additionally, the amount of groups used was half inspired by the literature [1, 2], which present tests on 3 - 7 groups initially, and half based on our own evidence, which stabilize results testing 5 groups.

5.2. Dataset

The dataset used in this work is MovieLens, which contains 100,000 ratings over 9,000 movies evaluated by 600 users [21].

Table 1: Data sample used in the experiment.

MovieId	Title	Genres
1	Toy Story	Adventure, Animation...
2	Jumanji	Adventure, Children...
3	Heat	Action, Crime

Some data processing is applied to remove the year information from the title description. Such a noise impacts negatively in the similarity calculation. The movie titles are genuinely in the format ***"Lion King, The (1994)"***. Also, we parse genres classification to fit in a list structure to ease data manipulation. Table 1 illustrates the pruned data used in the evaluation.

5.3. Metrics

5.3.1. Intra List Diversity (ILD)

5.3.1. Intra List Diversity (ILD) is a metric that compares how diverse items are in a list [22], and as the output shows the diversity score, from 0 to 1, where 0 means there is no diversity and 1 that the list is completely diverse. In other words, this metric evaluates how dissimilar the items in a list are. The ILD is inverse of the Intra List Similarity (ILS) metric [15], as in the ILD the distance between items is considered rather than similarity as in ILS. We implemented the ILD score in a list of items R as shown in Equation 7.

$$ILD(R) = \frac{\sum_{i \in R} \sum_{j \in R/\{i\}} Distance(i,j)}{2}, \quad (7)$$

where

$$Distance(i,j) = 1 - Similarity(i,j) \qquad (8)$$

Equation 7 calculates the mean of all distances from pairs *(i,j)* in list *R*, expect when *i==j*. The distance from a pair *(i,j)* is defined in Equation 8 as the complement of the similarity score for this pair.

5.3.2. Precision@K metric is the percentage of relevant items recommended to the user at k level in the recommended list [23], as shown in Equation 9.

$$P@k = \frac{RIK}{len(K)}, \qquad (9)$$

Where k is the top of the list to evaluate, RIK stands for **R**elevant **I**tems in the list at **K**, and $len(k)$ is the size of the list at k. For example, in a set of $k = 3$, if we have only 1 relevant item and it is in the first position, then $p@3 = \frac{1}{3}$. This metric compares the output items with a ground-truth. As groups are randomly generated, we adapted from experiments with real users to evaluate precision in the experiment. Inspired by [8], we set the relevant recommendations as those with a higher score than the global mean of evaluated movies. Also, it is important to highlight that experiments on randomly formed groups have not exceeded 0.4 precision score for a group of 5, as seen in the literature [24].

We generate groups profile based on some predictions, what leads to an uncertain in the ground-truth for groups. As groups are synthetically generated, we made an adaptation to evaluate precision on the experiment. We define the condition that the relevant recommendations are those with higher score than the global mean of evaluated movies [8].

5.4. Results

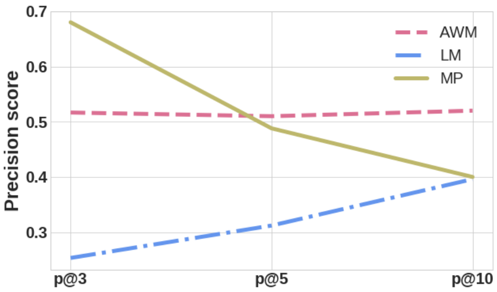


Figure 7: GD approach.

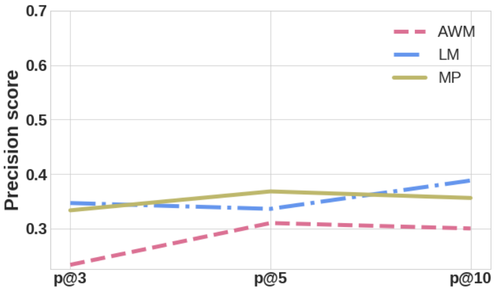


Figure 8: RD approach.

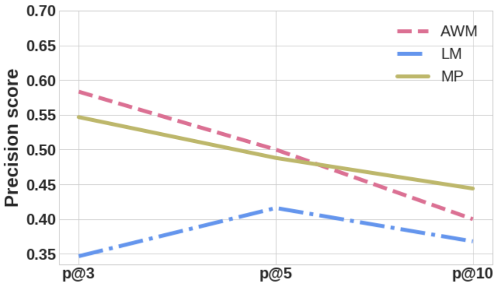


Figure 9: SD approach.

First we present the precision scores of aggregation techniques for the following approaches GD (Greedy), RD (Bounded Random), and SD (Standard). The overall best performance of aggregation techniques occurs in the GD as seen in Figure 7. The AWM score steadily increases over 0.5. Furthermore, the precision decreases from 0.68 (p@3) to 0.4 (p@10) when the MP technique is tested. The LM technique score increases, reaching 0.4. As to the RD approach, as seen in Figure 8, all three technique scores perform under 0.4 at p@3, p@5, and p@10. Figure 9 shows the SD approach with no diversification. Both AWM and MP scores start well at p@3 but then decline at p@10.

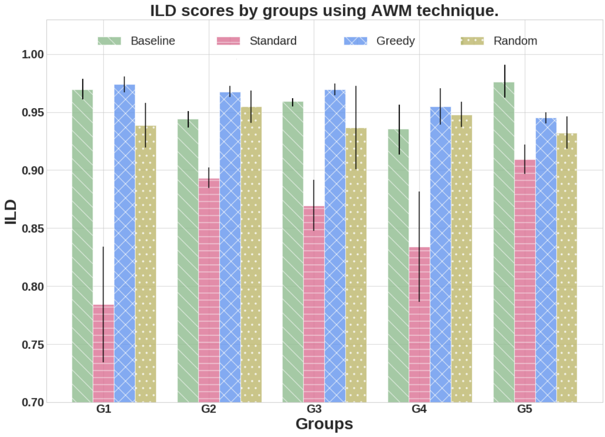


Figure 10: ILD results calculated for different groups and approaches.

Given the results obtained from the three techniques, we proceed to assess the diversity algorithms using the AWM technique, as it achieved the best results. Figure 12 exposes the precision scores using AWM aggregated by diversification approaches.

Figure 10 demonstrates the ILD scores obtained by four approaches in five different groups. The black line in the top of bars represents the standard deviation error (std). For the groups G2, G3, and G4, the GD approach performs better than BL even considering the std. For G1, both BL and GD approaches achieve similar results scoring 0.9694 and 0.9739 respectively. G5 is the only group where BL performs better than GD. The RD approach performs lower scores than the GD in all groups. It is also noticeable that RD presents a much higher std score than GD, achieving the top score of $8,5\%$ on std. The SD approach achieves the worst ILD results from all approaches. This outcome is somehow expected because this is the only approach that does not implement diversity. Overall, the GD approach achieves the closest result to the BL.

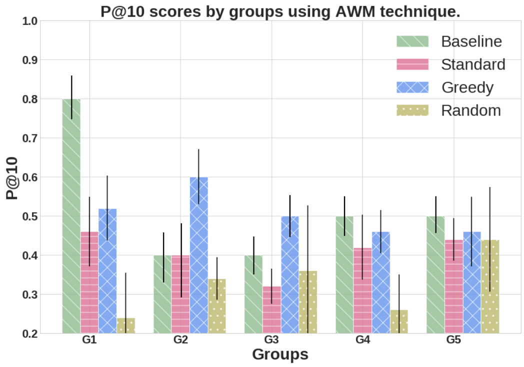


Figure 11: Precision at 10 calculated in different groups and approaches.

Figure 11 illustrates the Precision@10 scores obtained by the four approaches in five different groups. Analyzing the top-10 recommended items in each approach, the BL approach performs better in three groups: G1, G4, and G5, with a large advantage in G1 (approximately 20%). However, at G2, the GD approach shows a large advantage on BL, achieving 28% in advantage. The GD approach also performs better in G3 with a 10% advantage over BL. Furthermore, GD achieves the least std value among the tested approaches, except by BL. The latter achieved peaks of 8%, being a consistent behaviour in comparison to the other approaches.

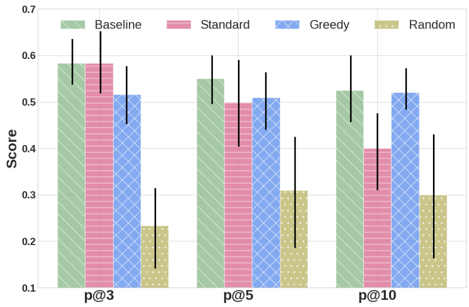


Figure 12: Precision scores aggregated by diversification approaches.

Figure 12 presents a panorama of the four diversification approaches on three levels of precision: precision at 3 (p@3), precision at 5 (p@5), and precision at 10 (p@10). The BL approach performs better in the three levels, however, it is equivalent to the SD approach at p@3 reaching 0.58, and it performs equally to GD at p@10 scoring 0.52. The GD approach remains constant with a slight increase from p@5 score to p@10 score, and it scored over 0.5 at all levels alongside BL. The SD approach falls steadily over the levels, achieving 0.4 at p@10. The RD approach achieves the worst results, it peaks at p@5 scoring 0.31 and it remains at the bottom of all levels evaluated.

As a result, we can observe that BL performs slightly worse than the GD approach as to the precision evaluation. On average, the GD approach performs 3.8% better than BL. Also, GD performs slightly better than BL as to the diversity evaluation, with an advantage of 1.8% on average. Even for a prediction approach, the GD performs consistently in both precision and diversity evaluation. In summary, the experimental evaluation indicates that the proposal using the GD approach is advised to be implemented alongside BL in a further online experiment with real users.

6. Conclusion

This paper presents a group recommendation model that suggest relevant items for groups based on a group profile and using diversification algorithms. The experiments using randomly formed groups evaluated precision and diversity of the recommendations. The results obtained with experiments indicate the best approach to implement in a further experiment with real users. As the GD approach performs slightly better than the BL approach, at least these two approaches will be presented in a further experiment. An important highlight is that this proposal was built with open source technologies and available at GitHub[1].

As to the future works, we plan to assess other

[1] `codebase:https://github.com/amandachagas/GRSwithDiversity.git`

group formation strategies despite random selection, such as grouping users by demographic information, geographic position, or personal interests. Furthermore, even though there is an increase of diversity in this model, it is possible to evaluate implicit factors such as fairness in genres distribution among items. Also, including more descriptive features in the dataset, such as synopsis or cast, can be subject to assessment. Additionally, a sensitivity analysis regarding the group size may present interesting outcomes as well as a more extensive numerical study including a greater number of groups. Moreover, applying the aggregation technique dynamically according to the group characteristics may enhance diversity as well. Finally, experiments with real users will certainly lead to important insights that could help improve the overall group recommendation model.

Acknowledgment

This work is supported by FAPESB BOL0596/2019.

References

[1] J. Masthoff, *Group Recommender Systems: Combining Individual Models*, pp. 677–702. Boston, MA: Springer US, 2011.

[2] S. Amer-Yahia, S. B. Roy, A. Chawlat, G. Das, and C. Yu, "Group recommendation: Semantics and efficiency," *Proc. VLDB Endow.*, vol. 2, pp. 754–765, Aug. 2009.

[3] A. Jameson and B. Smyth, *Recommendation to Groups*, pp. 596–627. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

[4] K. Bradley and B. Smyth, "Improving recommendation diversity," in *Proceedings of the Twelfth Irish Conference on Artificial Intelligence and Cognitive Science, Maynooth, Ireland*, pp. 85–94, Citeseer, 2001.

[5] M. Kaminskas and D. Bridge, "Diversity, serendipity, novelty, and coverage: A survey and empirical analysis of beyond-accuracy objectives in recommender systems," *ACM Trans. Interact. Intell. Syst.*, vol. 7, pp. 2:1–2:42, Dec. 2016.

[6] T. T. Nguyen, T. C. Phan, T. T. Nguyen, Q. V. H. Nguyen, and B. Stantic, "Diversifying group recommendation," *IEEE Access*, vol. 6, pp. 17776–17786, 2018.

[7] M. Atas, S. Reiterer, A. Felfernig, T. N. T. Tran, and M. Stettinger, "Polarization effects in group decisions," in *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*, UMAP '18, (New York, NY, USA), p. 305–310, Association for Computing Machinery, 2018.

[8] O. Kaššák, M. Kompan, and M. Bieliková, "Personalized hybrid recommendation for group of users: Top-n multimedia recommender," *Information Processing Management*, vol. 52, no. 3, pp. 459 – 477, 2016.

[9] A. Pujahari and D. S. Sisodia, "Aggregation of preference relations to enhance the ranking quality of collaborative filtering based group recommender system," *Expert Systems with Applications*, vol. 156, p. 113476, 2020.

[10] K. Tsukuda and M. Goto, "Dualdiv: Diversifying items and explanation styles in explainable hybrid recommendation," in *Proceedings of the 13th ACM Conference on Recommender Systems*, RecSys '19, (New York, NY, USA), p. 398–402, Association for Computing Machinery, 2019.

[11] M. Kaya and D. Bridge, "Community-aware diversification of recommendations," in *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, SAC '19, (New York, NY, USA), p. 1639–1646, Association for Computing Machinery, 2019.

[12] M. Kaya and D. Bridge, "Subprofile-aware diversification of recommendations," *User Modeling and User-Adapted Interaction*, vol. 29, no. 3, pp. 661–700, 2019.

[13] D. Bertram, "Likert scales," *Retrieved November*, vol. 2, p. 2013, 2007.

[14] B. Smyth and P. McClave, "Similarity vs. diversity," in *Case-Based Reasoning Research and Development* (D. W. Aha and I. Watson, eds.), (Berlin, Heidelberg), pp. 347–361, Springer Berlin Heidelberg, 2001.

[15] C.-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen, "Improving recommendation lists through topic diversification," in *Proceedings of the 14th International Conference on World Wide Web*, WWW '05, (New York, NY, USA), p. 22–32, ACM, 2005.

[16] A. G. Galliano, ed., *Introdução à sociologia*. Harper Row do Brasil, 1981.

[17] R. Ahuja, A. Solanki, and A. Nayyar, "Movie recommender system using k-means clustering and k-nearest neighbor," in *2019 9th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, pp. 263–268, 2019.

[18] S. Girase, D. Mukhopadhyay, *et al.*, "Role of matrix factorization model in collaborative filtering algorithm: A survey," *arXiv preprint arXiv:1503.07475*, 2015.

[19] P. Lops, M. de Gemmis, and G. Semeraro, "Content-based recommender systems: State of the art and trends," in *Recommender Systems Handbook* (F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, eds.), pp. 73–105, Springer US, 2011.

[20] G. Adomavicius, Tuzhilin, and Alexander, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, pp. 734–749, jun 2005.

[21] F. M. Harper and J. A. Konstan, "The movielens datasets: History and context," *ACM Trans. Interact. Intell. Syst.*, vol. 5, Dec. 2015.

[22] M. Rashidi, A. R. C. Hussin, and M. Nilashi, "Entropy-based ranking approach for enhancing diversity in tag-based community recommendation," *Journal of Soft Computing and Decision Support Systems*, vol. 3, no. 1, pp. 1–7, 2015.

[23] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. New York, NY, USA: Cambridge University Press, 2008.

[24] L. Quijano-Sánchez, J. A. Recio-García, and B. Díaz-Agudo, "Group recommendation methods for social network environments," in *3rd workshop on recommender systems and the social web within the 5th ACM international conference on recommender systems (RecSys' 11)*, pp. 24–31, 2011.