Group Recommendation of Board Games Using Aggregation Strategies

Mathias Lambert V. [1 2] **Jaime Pérez** [1]

Abstract

This work addresses the challenge of recommending board games to groups of users by implementing and evaluating a group recommendation system. Using a dataset from BoardGameGeek, random groups with similar preferences were formed, and four individual recommendation models were trained: Random, Most Popular, SVD++, and an XGB Regressor. Subsequently, various aggregation strategies were applied to generate group recommendations. Results indicate that the SVD++ model offers superior performance for predicting individual ratings. However, when forming recommendation lists for groups, the combination of the XGB Regressor with strategies such as "average" and "most pleasure" achieved the highest performance in terms of the NDCG@10 metric, demonstrating the effectiveness of combining robust individual predictions with suitable aggregation methods for this social context.

1. Introduction

Most work on recommender systems focuses on recommending items to individual users. For example, they might recommend a book to a specific user based on a model of their past preferences. In this case, preferences can be implicit (e.g., clicks, purchases, viewing time, etc.) or explicit (e.g., likes, ratings, rankings, etc.) (Ricci et al., 2022).

There are many situations where it would be beneficial to recommend to a group of users rather than an individual. For example, a recommender system could select TV shows for a group to watch or a sequence of songs to listen to, based on the individual preferences of all group members. Recommending to groups is more complicated than recommending to individuals. Assuming we know perfectly what is good for individual users, the question arises of how to define and find what is also good for a group composed of the same individuals. The usual approach to solving this question often revolves around combining individual user preferences into a group preference model, or tailoring individual recommendations into group recommendations. Methods or algorithms that combine individual user preferences or recommendations are called preference aggregation strategies (Masthoff & Delić, 2022).

Table 1. Example of individual preferences and *Least Misery Strategy*

User\Item	A	B	C	D	E	F	G	H	I	J
Pedro	10	4	3	6	10	9	6	8	10	8
Juan	1	9	8	9	7	9	6	9	3	8
Diego	10	5	2	7	8	8	5	6	7	6
Least Misery Strategy	1	4	2	6	7	8	5	6	3	6

The main problem that group recommendation must solve is how to adapt to the group as a whole, based on information about individual users' likes and dislikes. For example, suppose the group consists of three people: Pedro, Juan, and Diego. Assume a system is aware that these three persons are present and knows their interests regarding a set of items (e.g., music clips or advertisements). Table 1 provides examples of scores on a scale from 1 (hates it very much) to 10 (likes it very much). Which items should the system recommend?

Many strategies exist for aggregating individual scores into a group rating. For example, the Least Misery Strategy uses the minimum score to avoid dissatisfaction among group members (Table 1). These strategies can then be used to rank items according to their relevance for the group in question, and thus estimate what should be recommended to the group.

Some group aggregation techniques are presented in Table 2. More complex techniques involve using heuristics, meaning they consider assumptions about decision-making. For example, they might identify that certain users don't give high ratings despite approving of or liking an item, and these assumptions are then incorporated into the aggregation strategy's design. For instance, Spearman foot-rule rank aggregation (Baltrunas et al., 2010) states that the ag-

[1]Pontificia Universidad Catolica de Chile, Santiago, Chile [2]Biomedical Imaging Center, Pontificia Universidad Catolica de Chile, Santiago, Chile. Correspondence to: Mathias Lambert V. <mglambert@uc.cl>.

Proceedings of the 42^{nd} International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

Table 2. Aggregation strategies.

Strategy	How it works	Example based on Table 1
Plurality/majority voting	Uses 'first past the post':repetitively, the item with the most votes is chosen.	A is chosen first, as it has the highest rating for the majority of the group, followed by E (which has the highest rating for the majority when excluding A)
Average	Averages individual ratings.	B's group rating is 6, namely (4+9+5)/3
Multiplicative	Multiplies individual ratings.	B's group rating is 180, namely 4*9*5
Borda count	Counts points from items' rankings in the individuals' preference lists, with bottom item getting 0 points, next one up getting one point, etc.	A's group rating is 17, namely 0 (ast for Juan) + 9 (first for Diego) + 8 (shared top 3 for Pedro)
Copeland score	Counts how often an item beats other items (using majority vote) minus how often it looses.	F's group rating is 5, as F beats 7 items (B,C,D,G,H,I,J) and looses from 2 (A,E)
Approval voting	Counts the individuals with ratings for the item above approval threshold (e.g. 6).	B's group rating is 1 and F's is 3
Least Misery	Takes the minimum of individual ratings.	B's group rating is 4, namely the smallest of 4,9,5
Most pleasure	Takes the maximum of individual ratings.	B's group rating is 9, namely the largest of 4,9,5
Average without Misery	Averages individual ratings, after excluding items with individual ratings below a certain threshold (say 4).	J's group rating is 7.3 (the average of 8,8,6), while A is excluded because Juan hates it
Fairness	Items are ranked as if individuals are choosing them in turn.	Item E may be chosen first (highest for Pedro), followed by F (highest for Juan) and A (highest for Diego)
Most respected person (or Dictatorship)	Uses the rating of the most respected individual.	If Juan is the most respected person, then A's group rating is 1. If Diego is most respected, then it is 10

gregated list for the group is the one with the minimum distance from the individual lists. The Spearman foot-rule distance between two lists is defined as the sum of the absolute differences between the positions of items in those lists.

"Influence-based methods" are a type of aggregation strategy that incorporate the idea that the opinions of group members can influence each other when forming a consensus. Essentially, these methods quantify how much each individual values the opinions of others and use this measure to adjust initial preferences. Examples include the Pre-GROD and GROD methods (Castro et al., 2018). These models begin by creating a "trust matrix" that reflects the importance each group member assigns to the opinions of others. Based on the premise that opinions evolve under mutual influence, individual ratings are iteratively updated using the trust matrix until a consensus is reached. At that point, the group's rating for each item is calculated as the average of the updated opinions. A distinctive feature of GROD is its ability to ensure consensus; if it's not reached spontaneously, the method adjusts the trust matrix by adding elements that facilitate the convergence of opinions.

The After Factorization (AF), Before Factorization (BF), and Weighted Before Factorization (WBF) methods (Ortega et al., 2016) employ Matrix Factorization (MF), with the only difference being whether individual preferences are aggregated (using standard aggregation strategies) before or after MF. The interesting approach here is WBF, which aggregates individual user ratings but assigns more weight to items that have been rated by the majority of the group and have similar ratings from the group's users.

The AGREE method (Cao et al., 2018) uses a neural network to learn the aggregation strategy and item-dependent weights of each group member. The network extends the Neural Collaborative Filtering (NCF) structure (He et al., 2017) by learning not only user-item interactions but also group-item interactions.

MoSAN (Vinh Tran et al., 2019) is another neural network method that models user-user interactions between members with the help of attention blocks, for ad-hoc groups that had no prior interactions. For each group, a set of sub-attention networks is created (one for each group member). The network aims to capture each group member's decisions, given the decisions of others in that group.

Finally, GroupIM (Sankar et al., 2020) trains a neural network to calculate the probabilities of a group interacting with an item by minimizing the group recommendation loss between the group-item interaction history and the predicted item probabilities, and the contextually weighted user-item loss.

The evaluation of group recommendation systems (Felfernig et al., 2018) uses methodologies similar to those employed in systems for individual users, differing primarily in offline and online evaluation protocols. Offline evaluation uses datasets with user ratings to estimate the predictive quality of algorithms, dividing data into training and test sets to apply classification metrics (precision and recall), error metrics (mean absolute error), and ranking metrics (discounted cumulative gain). Meanwhile, online evaluation is conducted through user studies, either in laboratories under controlled experimental designs, via crowd-sourcing platforms, or in production environments with A/B tests. In addition to these predictive performance metrics, there are specific metrics for the group context that go beyond accuracy, such as the level of consensus among members, fairness in satisfying individual preferences, coverage of the recommended item catalog, and serendipity, or the system's ability to offer novel and unexpected recommendations.

Considering that board games constitute an inherently social activity, the decision to purchase a new title benefits from taking into account the preferences of all members of the gaming group. This particularity poses the challenge of developing a system that recommends products not for an individual, but for a set of users simultaneously. Therefore, this work implements a group recommendation method designed for selecting board games and compares its performance with individual recommendation strategies to validate its usefulness.

2. Dataset

The dataset used in this study, called Board Game Database from BoardGameGeek (Wadkins, 2022), was obtained from Kaggle and represents a complete extraction of publicly available information from BoardGameGeek (BGG), one of the largest collaborative board game databases.

BGG dataset originally contains:

- 21,925 games ranked in 8 categories plus a general category.
- 411,375 users.
- 18,942,216 ratings.

In order to reduce the computational cost, we select the 50 best ranked games for each category and the user with more than 50 reviews. Resulting in 275 games and 8,231 users.

2.1. Groups' Formation

BGG is a dataset that does not contain group information. To form the groups, given the characteristics of the board game context, we use a random groups method (Carvalho & Macedo, 2013), ideal for a number of people sharing the same environment at a given time, without a clear motivation to be linked.

Specifically, for this article, we utilized the Random Similar Groups (RSG) method (Ghaemmaghami & Salehi-Abari, 2021), which randomly selects l groups from a preference dataset. In this method, the group size is randomly picked from $[s_{min}, s_{max}]$, and group members are required to have similar preferences, enforced by ensuring all pairwise Kendall-τ correlations of group members are at least τ_{sim}. The parameters used were $s_{min} = 2$, $s_{max} = 10$, and $\tau_{sim} = 0.5$.

2.2. Pre-processing

For each board game, the description field and the image of its cover were used. For the description and the image, their embedding vectors were obtained using BERT(Devlin et al., 2018) and ViT(Wu et al., 2020) respectively. To reduce the computational cost, a dimensionality reduction using 15-component PCA was applied separately to both vectors.

We created 247 consistent groups, i.e. the members have at least 2 games in common. For each group, the items in common were set aside as test items while the rest were left for training.

3. Methods

The experimental method consisted of training the following individual recommendation models:

1. Random recommender.
2. Most popular recommender.
3. SVD++. (Koren, 2008; 2010; Hug, 2020)
4. XGB regressor. (Chen & Guestrin, 2016)

With these models we evaluate the individual rating estimate and evaluate the average performance within each group between the aggregate recommendation list and the individual preference list using the methods in table 2 (except for Plurality).

3.1. XGB regressor

This model is based on a regression approach. Its inputs include a user representation vector, computed as the rating-weighted average of the embedding vectors of the games the user has rated. Additionally, it incorporates the cosine distances between the user's description and image components and those of the candidate games.

4. Results

4.1. Individual recommendations

Table 3. Individual rating prediction results

Metric	Random	Most popular	SVD++	XGB
MAE	2.8284	2.1885	0.7807	0.9064
RMSE	3.4602	2.4680	1.0414	1.1814

4.2. Group list

Table 4. Precision@10 (mean ± str)

Aggregation method	Most popular	Random	SVD++	XGB
approval_voting	0.402 ± 0.169	0.160 ± 0.077	0.156 ± 0.083	0.074 ± 0.044
average	0.402 ± 0.169	0.158 ± 0.083	0.156 ± 0.083	0.067 ± 0.029
average_without_misery	0.246 ± 0.114	0.090 ± 0.022	0.118 ± 0.040	0.000 ± 0.000
borda_count	0.402 ± 0.169	0.158 ± 0.083	0.156 ± 0.083	0.058 ± 0.028
copeland_score	0.402 ± 0.169	0.165 ± 0.098	0.156 ± 0.083	0.072 ± 0.032
fairness	0.402 ± 0.169	0.158 ± 0.083	0.156 ± 0.083	0.061 ± 0.030
least_misery	0.402 ± 0.169	0.156 ± 0.080	0.156 ± 0.083	0.073 ± 0.043
most_pleasure	0.402 ± 0.169	0.159 ± 0.083	0.156 ± 0.083	0.070 ± 0.027
most_respected_person	0.402 ± 0.169	0.167 ± 0.096	0.156 ± 0.083	0.073 ± 0.031
multiplicative	0.402 ± 0.169	0.156 ± 0.080	0.156 ± 0.083	0.067 ± 0.029

Table 5. NDCG@10 (mean ± str)

Aggregation method	Most popular	Random	SVD++	XGB
approval_voting	0.741 ± 0.136	0.796 ± 0.170	0.861 ± 0.158	0.980 ± 0.059
average	0.741 ± 0.136	0.799 ± 0.169	0.849 ± 0.168	0.993 ± 0.035
average_without_misery	0.817 ± 0.161	0.130 ± 0.336	0.081 ± 0.273	0.324 ± 0.469
borda_count	0.741 ± 0.136	0.797 ± 0.170	0.858 ± 0.159	0.975 ± 0.064
copeland_score	0.741 ± 0.136	0.788 ± 0.167	0.858 ± 0.159	0.983 ± 0.053
fairness	0.741 ± 0.136	0.796 ± 0.168	0.852 ± 0.161	0.976 ± 0.063
least_misery	0.741 ± 0.136	0.811 ± 0.173	0.858 ± 0.159	0.979 ± 0.058
most_pleasure	0.741 ± 0.136	0.797 ± 0.169	0.843 ± 0.169	0.994 ± 0.034
most_respected_person	0.741 ± 0.136	0.789 ± 0.168	0.856 ± 0.159	0.983 ± 0.053
multiplicative	0.741 ± 0.136	0.811 ± 0.173	0.858 ± 0.159	0.993 ± 0.035

5. Discussion

For individual recommendations, the results show a clear superiority of the SVD++ model, which obtains MAE and RMSE errors considerably lower than the XGBoost regressor and baseline models. This suggests that for predicting a user's exact rating on this dataset, SVD++'s collaborative filtering approach is more effective than the content-based feature approach used by the XGB model.

In the context of group recommendations, a highly interesting phenomenon is observed. Despite SVD++'s superior individual performance, the XGB regressor, when combined with aggregation strategies, consistently produces better results in the NDCG@10 ranking metric. Specifically, the ”average” and ”most pleasure” strategies achieve the highest scores. This indicates that while XGB is less precise at predicting exact ratings, its ability to correctly rank items preferred by each individual translates into a higher-quality recommendation list for the group once preferences are aggregated. The effectiveness of the ”most pleasure” strategy suggests that for a social activity like board games, maximizing the satisfaction of the most enthusiastic group member is a very effective approach.

6. Conclusions

This work successfully implemented and evaluated a group recommendation system for board games. It concludes that for predicting individual ratings, the SVD++ model is the most accurate and effective, as demonstrated by the error metrics.

However, the main finding is that the best individual prediction model doesn't necessarily yield the best group recommendations. For the final task of generating a recommendation list for a group, the combination of the XGB Regressor model with ”average” and ”most pleasure” aggregation strategies proved most effective, achieving the highest performance in the NDCG@10 metric. These results validate the utility of group recommendation approaches for social activities and underscore the importance of optimizing for the quality of the final ranking over the exactness of rating predictions.

Software

All codes used in this article are available in https://github.com/mglambert/RecSysProject.

Authors contributions

- Mathias Lambert: Exploratory data analysis for the initial submission, research on aggregation methods, programming of all code available in the repository, nearly complete writing of the progress report, complete writing of the poster, and complete writing of this article.

- Jaime Perez: Writing of the initial submission, baselines for the initial submission, partial implementation of a method in the progress report. Writing of a parallel article to this one with the erroneous results presented at the poster session (submission 1).

References

Baltrunas, L., Makcinskas, T., and Ricci, F. Group recommendations with rank aggregation and collaborative filtering. In *Proceedings of the Fourth ACM Conference on Recommender Systems*, RecSys '10, pp. 119–126, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781605589060. doi: 10.1145/1864708.1864733. URL https://doi.org/10.1145/1864708.1864733.

Cao, D., He, X., Miao, L., An, Y., Yang, C., and Hong, R. Attentive group recommendation. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '18, pp. 645–654, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450356572. doi: 10.1145/3209978.3209998. URL https://doi.org/10.1145/3209978.3209998.

Carvalho, L. A. M. and Macedo, H. T. Generation of coalition structures to provide proper groups' formation in group recommender systems. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13 Companion, pp. 945–950, New York, NY, USA, 2013. Association for Computing Machinery. ISBN 9781450320382. doi: 10.1145/2487788.2488089. URL https://doi-org.pucdechile.idm.oclc.org/10.1145/2487788.2488089.

Castro, J., Lu, J., Zhang, G., Dong, Y., and Martínez, L. Opinion dynamics-based group recommender systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 48(12):2394–2406, 2018. doi: 10.1109/TSMC.2017.2695158.

Chen, T. and Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pp. 785–794, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450342322. doi: 10.1145/2939672.2939785. URL https://doi.org/10.1145/2939672.2939785.

Devlin, J., Chang, M., Lee, K., and Toutanova, K. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805, 2018. URL http://arxiv.org/abs/1810.04805.

Felfernig, A., Boratto, L., Stettinger, M., and Tkalčič, M. *Evaluating Group Recommender Systems*, pp. 59–71. Springer International Publishing, Cham, 2018. ISBN 978-3-319-75067-5. doi: 10.1007/978-3-319-75067-5_3. URL https://doi.org/10.1007/978-3-319-75067-5_3.

Ghaemmaghami, S. S. and Salehi-Abari, A. Deepgroup: Representation learning for group recommendation with implicit feedback, 2021. URL https://arxiv.org/abs/2103.07597.

He, X., Liao, L., Zhang, H., Nie, L., Hu, X., and Chua, T.-S. Neural collaborative filtering, 2017. URL https://arxiv.org/abs/1708.05031.

Hug, N. Surprise: A python library for recommender systems. *Journal of Open Source Software*, 5(52):2174, 2020. doi: 10.21105/joss.02174. URL https://doi.org/10.21105/joss.02174.

Koren, Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08, pp. 426–434, New York, NY, USA, 2008. Association for Computing Machinery. ISBN 9781605581934. doi: 10.1145/1401890.1401944. URL https://doi.org/10.1145/1401890.1401944.

Koren, Y. Factor in the neighbors: Scalable and accurate collaborative filtering. *ACM Trans. Knowl. Discov. Data*, 4(1), January 2010. ISSN 1556-4681. doi: 10.1145/1644873.1644874. URL https://doi.org/10.1145/1644873.1644874.

Masthoff, J. and Delić, A. *Group Recommender Systems: Beyond Preference Aggregation*, pp. 381–420. Springer US, New York, NY, 2022. ISBN 978-1-0716-2197-4. doi: 10.1007/978-1-0716-2197-4_10. URL https://doi.org/10.1007/978-1-0716-2197-4_10.

Ortega, F., Hernando, A., Bobadilla, J., and Kang, J. H. Recommending items to group of users using matrix factorization based collaborative filtering. *Information Sciences*, 345:313–324, 2016. ISSN 0020-0255. doi: https://doi.org/10.1016/j.ins.2016.01.083. URL https://www.sciencedirect.com/science/article/pii/S0020025516300196.

Ricci, F., Rokach, L., and Shapira, B. *Recommender Systems: Techniques, Applications, and Challenges*, pp. 1–35. Springer US, New York, NY, 2022. ISBN 978-1-0716-2197-4. doi: 10.1007/978-1-0716-2197-4_1. URL https://doi.org/10.1007/978-1-0716-2197-4_1.

Sankar, A., Wu, Y., Wu, Y., Zhang, W., Yang, H., and Sundaram, H. Groupim: A mutual information maximization framework for neural group recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '20, pp. 1279–1288, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450380164. doi: 10.1145/3397271.3401116. URL https://doi.org/10.1145/3397271.3401116.

Vinh Tran, L., Nguyen Pham, T.-A., Tay, Y., Liu, Y., Cong, G., and Li, X. Interact and decide: Medley of sub-attention networks for effective group recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR'19, pp. 255–264, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450361729. doi: 10.1145/3331184.3331251. URL https://doi.org/10.1145/3331184.3331251.

Wadkins, J. Board game database from boardgamegeek, 2022. URL https://www.kaggle.com/datasets/threnjen/board-games-database-from-boardgamegeek/data.

Wu, B., Xu, C., Dai, X., Wan, A., Zhang, P., Yan, Z., Tomizuka, M., Gonzalez, J., Keutzer, K., and Vajda, P. Visual transformers: Token-based image representation and processing for computer vision, 2020.