Analyzing the Impact of Acoustic Features on Music Recommendation for Children Across Age Groups

Erkin Başol

Responsible Professor: Sole Pera Supervisor: Robin Ungruh TU Delft CSE3000 Research Project

e.basol-1@student.tudelft.nl https://github.com/Protestak/RP_Project

Introduction

- Collaborative filtering is a widely used algorithm in music recommender systems [1].
- Music Recommender Systems are generally targeted towards adults [2].
- Children have distinct music preferences and needs [2].
- Despite their diverse preferences, there is a lack of recommender systems designed for children that can optimally serve them.
- Acoustic features influence children's music preferences [3]; however, recommenders that employ individual acoustic features are lacking.
- The goal is to determine which acoustic features most enhance the performance when incorporated into a Collaborative Filtering algorithm for children.

Research Question

To what degree can the incorporation of different acoustic features improve the performance of a CF-based recommender system for children?

Experimental Setup

- Top-10 offline recommendation experiments on users aged 15–18 (279 users per group).
- 11 recommenders trained: 1 baseline + 10 extended with individual acoustic feature values, extending item-KNN by replacing binary interaction values with acoustic feature content values.
- Compare extended models to baseline on ranking and diversity metrics to find features that most improve performance that could be employed in child-centric recommenders.

Data Preprocessing

- Use LFM-2B dataset [4] for listening events of children, and LFM-BeyMS dataset [5] for acoustic track features.
- Use listening events with complete track features for users aged 12–18.
- Focus on the year 2012; split data into Training (Jan–May), Validation (Jun–Jul), and Test (Aug–Oct) sets.
- Preprocess to maximize the number of users while maintaining sufficient interactions to reduce sparsity, and balance age groups by retaining only those from 15 to 18.
- Final dataset contains 1116 users (279 per age group), with an average of 11.62 interactions in the test set, 10.50 in the validation set, and 38.63 in the training set, across 4700 distinct songs.

Evaluation Metrics

- Mean Reciprocal Rank – First hit
- Normalized Discounted Cumulative Gain – Rank quality
- Hit Rate – Hit Probability
- Intra-List Diversity – Item variety
- Catalog Coverage – Catalog reach

References

[1] M. Schedl, P. Knees, B. McFee, and D. Bogdanov, *Music Recommendation Systems: Techniques, Use Cases, and Challenges*. New York, NY: Springer US, 2022, pp. 927–971. [Online]. Available: https://doi.org/10.1007/978-1-0716-2197-4_24

[2] L. Spear, A. Milton, G. Allen, A. Raj, M. Green, M. D. Ekstrand, and M. S. Pera, "Baby shark to barracuda: Analyzing children's music listening behavior," in *Proceedings of the 15th ACM Conference on Recommender Systems*, ser. RecSys '21. New York, NY, USA: Association for Computing Machinery, 2021, pp. 639–644. [Online]. Available: https://doi.org/10.1145/3460231.3478856

[3] I. Papadimitriou, "Leveraging children's music preferences to enhance the recommendation process," Master's thesis, Delft University of Technology, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft, Netherlands, 2024. [Online]. Available: https://repository.tudelft.nl/file/File_a38a6a7a-2fd3-4a89-b58e-e905a4ec3a8f

[4] R. Ungruh, A. Bellogín, and M. S. Pera, "The impact of mainstream-driven algorithms on recommendations for children," in *Advances in Information Retrieval*, C. Hauff, C. Macdonald, D. Jannach, G. Kazai, F. M. Nardini, F. Pinelli, F. Silvestri, and N. Tonellotto, Eds. Cham: Springer Nature Switzerland, 2025, pp. 67–84.

[5] P. Müllner, D. Kowald, M. Schedl, C. Bauer, E. Zangerle, and E. Lex, "Lfm-beyms," May 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3784765

[6] G. Carraturo, V. Pando-Naude, M. Costa, P. Vuust, L. Bonetti, and E. Brattico, "The major–minor mode dichotomy in music perception: A systematic review on its behavioural, physiological, and clinical correlates," *bioRxiv*, 2023. [Online]. Available: https://www.biorxiv.org/content/early/2023/03/18/2023.03.16.532764

Results and Discussion

Model	HitRate@10	MRR@10	NDCG@10	Coverage@10
Age Group 15/16				
Item-KNN (Baseline Model)	0.1326/0.1147	0.0361/0.0320	0.0187/0.0182	0.3503/0.3569
Item-KNN+Acousticness	0.1326/0.1183	**0.0402**/0.0306	0.0200/0.0190	0.3448/0.3567
Item-KNN+Instrumentalness	0.1577/0.1183	0.0416/0.0382	**0.0226**/0.0180	0.3280/0.3202
Item-KNN+Loudness	0.1470/0.1183	**0.0390**/0.0318	**0.0200**/0.0180	0.3503/0.3531
Item-KNN+Mode	0.1613/0.1039	0.0418/0.0324	0.0219/0.0162	0.3076/0.3127
Age Group 17/18				
Item-KNN (Baseline Model)	0.0968/0.1075	0.0181/0.0302	0.0099/0.0163	0.3756/0.3714
Item-KNN+Acousticness	0.1039/0.1219	0.0203/0.0315	**0.0116**/0.0168	0.3773/0.3695
Item-KNN+Instrumentalness	0.1111/0.1183	0.0276/0.0315	0.0123/0.0167	0.3416/0.3446
Item-KNN+Mode	**0.1434**/0.1398	**0.0390**/0.0387	**0.0169**/0.0196	0.3208/0.3263

Table 1. Performance Metrics by Age Group. Significant improvements over the baseline are bolded.

- **Mode** is the most prominent and impactful feature, consistently improving recommendations for users aged 15 and 17.
- **Mode** strongly influences emotions, with major modes evoking happiness and minor modes suggesting melancholy [6].
- For the ages 16 and 18, no feature significantly outperformed the baseline, indicating less consistent preferences.
- **Instrumentalness** and **acousticness** boost performance for age 15, with instrumentalness improving NDCG and acousticness enhancing MRR.
- **Mode** and **instrumentalness** reduce item coverage while significantly improving performance, highlighting a trade-off between accuracy and diversity, whereas features like **acousticness** and **loudness** maintain coverage and support more balanced recommendations.
- **Loudness** significantly improves recommendations for age 15, likely due to a preference for energetic music [2].

Responsible Research

- **Ethical research:** Used anonymized, historical datasets to avoid direct interaction with minors and ensure privacy compliance.
- **Reproducibility:** Open-source code with full pipeline, parameters, and documentation is provided on GitHub.

Conclusion

- Extending CF-based recommender systems with mode, instrumentalness, and acousticness can optimize the performance of recommenders for children and improve the recommendations quality.
- This contribution suggests that these features can improve an already widely used CF-based recommender, enabling it to better serve children.

Limitations & Future Work

- **Limited sample:** Needs more users for generalizability.
- **Outdated data:** Preferences in 2012 may not reflect current trends.
- **Implicit feedback:** No explicit ratings in the dataset to confirm intent.
- **Live testing:** Test these extended models in real-world systems using online evaluation.
- **Recent data:** Collect newer, rating-based datasets for accuracy.