Towards Personalizing Classical Music Recommendations

Markus Schedl
Department of Computational Perception
Johannes Kepler University Linz, Austria
Email: markus.schedl@jku.at

***Abstract*—While fans of classical music were found to be underrepresented on social media and music streaming platforms, they constitute an important target group for music recommender systems. We therefore focus on this group of listeners and investigate a wide range of recommendation approaches and variants for the task of music artist recommendation. Within the group of classical music listeners, we further assess categorizing users according to demographics and temporal music consumption behavior. We report the results of preliminary recommendation experiments and insights gained for the listener group under consideration.**

I. Introduction and Context

Music recommender systems have become a popular topic in research and business during the last few years [3], [9]. While tailoring recommendations according to certain user characteristics and contextual factors has been found important as well [1], [13], work on user-centric recommendation in the music domain is still in its infancy [8]. Existing work typically focuses on particular scenarios, for instance, music recommendation in cars [2] or for places of interest [7]. Exploited features include time, location, music descriptors, popularity trends [4], user activity [14], and user characteristics derived from listening profiles, such as mainstreaminess or taste diversity [6].

In contrast to previous work, we target a particular group of listeners, namely fans of classical music. This group constitutes a challenging audience for recommendation since respective listeners are underrepresented on social media and music streaming platforms [11]. Given the purchasing power of classical music aficionados, they nevertheless represent a crucial audience for music recommendation businesses. Furthermore, the PHENICX project,[1] in the context of which this study is conducted, aims at attracting new audiences for classical music, among others by exploiting social media.

II. Recommendation Approaches

We analyze stand-alone and hybrid recommendation approaches. Each user u has a listening profile L_u, which contains all items (artists) listened to. Furthermore, u is assigned a normalized playcount vector $\vec{p}_u$ containing the number of listening events over all items I. We consider the following **stand-alone approaches**:

PB: A popularity-based recommender that returns the N items listened to most frequently by all users.

CF: A user-based collaborative filtering approach that recommends N items listened to by u's nearest neighbors in terms of listening histories; neighbors are identified by computing the Inner product between $\vec{p}_u$ and $\vec{p}_v$, for every other user v.

LB: An extension to **CF** in that we consider as nearest neighbors of u only users that are located in the same country as u.

IB: A content-based (instance-based) approach where each item i is assigned a TF·IDF vector $\vec{w}_i$ in a vector space of *Last.fm* tags, in which tag weights are interpreted as term frequencies. The recommendation approach then identifies for each training item in L_u its nearest neighbors via maximizing cosine similarity between $\vec{w}_i$ and $\vec{w}_j \ \forall\, j \in J$, where J is the set of all items excluding i.

RB: A baseline that randomly picks users and recommends N artists they listened to.

For **CF**, **IB**, and **LB**, we investigate two *aggregation functions* to fuse the recommendations contributed by the nearest neighbors,[2] in case they are overlapping: arithmetic mean or maximum between the similarity scores of each item that is recommended by more than 1 neighbor.

In addition to the 8 stand-alone systems, we investigate a total of 192 algorithmic variants of **hybrid approaches**, which integrate combinations of **PB**, **CF**, **LB**, and **IB**. For these hybrid systems, we consider a variety of score *normalization functions* (n) and *fusion functions* (f): n_{none} (no normalization), n_{gauss} (Gaussian normalization), n_{sumto1} and n_{maxto1} (linear stretching of scores so that their sum or maximum equals 1, respectively); f_{max}, f_{mean}, f_{sum}, $f_{multiply}$ (fusing the scores of individual recommenders by computing their maximum, mean, sum, or product, respectively) and f_{borda} (rank aggregation based on Borda count [5], which has already proven successful for multimodal music recommendation [7]).

III. Listener Characteristics

In these preliminary experiments, we consider 3 aspects according to which we further group the fans of classical music under consideration: *age* (5 user groups in different age ranges), *country* (4 user groups according to the top countries in the dataset, i.e., USA, UK, Russia, and Germany), and *time of day* (3 user groups according to the time of day into which most of their listening activity falls: morning [7.00h–11.59h], afternoon [12.00h–19.59h], and night [20.00h–6.59h]).

[1] http://phenicx.upf.edu/

[2] In case of **CF** and **LB**, nearest neighbors refer to users; in case of **IB**, they refer to items.

IV. Experiments and Results

We conduct recommendation experiments on a subset of the dataset presented in [12]. The full dataset covers almost 200 million listening events by about 16,500 *Last.fm* users, who listen to more than 1 million unique artists. Since the work at hand focuses on fans of classical music, we filter users with less than 10% of listening events being to classical music, which yields a set of 362 listeners. For each combination of recommendation algorithm and user group, we perform 5-fold cross-validation on a per-user basis, i.e., we run for each user 5 experiments, iterating through all permutations of 80% unique training artists and 20% unique test artists. We compute average precision, recall, and F1-score as performance measures, averaged over all users and different numbers of recommended artists (10, 50, and 100).

Table I shows for all user categories under consideration the best performing approaches as well as the random baseline. These results can be interpreted as rules that allow us to tailor a recommendation method to a given subgroup of classical music-affine listeners. Summarizing the main findings, we observe that (i) all investigated recommendation approaches outperform the random baseline, (ii) hybrid methods tend to outperform stand-alone systems, (iii) the **PB** recommender performs best for teenagers, listeners in their mid to late twenties, users from Germany, and "afternoon listeners," (iv) hybrids of **CF** and **IB** frequently perform best in terms of recall, and (v) hybrids including **PB** and **LB** often perform best in terms of precision or F-score. Compared to our previous analysis [10], in which we used the same dataset, but did not restrict users to fans of classical music, performance in terms of F-score is consistently lower in the experiments at hand for all user groups and approaches.

TABLE I. Results (%) for best performing recommendation approaches for user aspects *age* (top), *country* (mid), and *time of day* (bottom).

Approach	Prec.	Rec.	F-score
Age: 6–17 (9)			
$RB(n_{none})$	**5.62**	6.97	4.84
$PB(n_{none})$	5.44	**8.34**	**5.00**
Age: 18–21 (57)			
$RB(n_{none})$	2.23	2.48	1.89
$PB + CF_{mean}(n_{maxto1}, f_{mean})$	5.57	**6.87**	4.50
$PB + CF_{mean} + LB_{mean}(n_{sumto1}, f_{mean})$	**5.68**	6.85	**4.61**
Age: 22–25 (86)			
$RB(n_{none})$	1.90	1.50	1.26
$CF_{mean} + IB_{max}(n_{none}, f_{max})$	**6.65**	**6.98**	**4.75**
Age: 26–30 (42)			
$RB(n_{none})$	1.69	1.19	1.11
$PB(n_{none})$	**6.48**	**4.51**	**3.95**
Age: 31–40 (29)			
$RB(n_{none})$	2.66	2.05	1.68
$PB(n_{none})$	6.05	**5.98**	4.15
$PB + LB_{mean}(n_{gauss}, f_{max})$	**6.13**	5.96	4.13
$PB + LB_{mean}(n_{maxto1}, f_{mean})$	6.12	5.98	**4.15**
Country: US (16)			
$RB(n_{none})$	3.91	2.97	2.56
$PB(n_{none})$	5.62	**6.31**	**4.28**
$CF_{mean} + IB_{max}(n_{none}, f_{max})$	**5.64**	5.43	4.12
Country: UK (14)			
$RB(n_{none})$	4.66	2.26	2.38
$PB(n_{none})$	6.62	**8.10**	4.56
$LB_{mean}(n_{none})$	**7.74**	4.04	4.15
$PB + LB_{mean}(n_{maxto1}, f_{mean})$	7.16	8.05	**4.62**
Country: RU (23)			
$RB(n_{none})$	2.98	1.61	1.71
$PB(n_{none})$	**6.71**	4.72	**4.37**
$CF_{mean} + IB_{max}(n_{none}, f_{max})$	5.94	**4.95**	4.24
Country: DE (20)			
$RB(n_{none})$	2.34	1.69	1.42
$PB(n_{none})$	**4.60**	**5.27**	**3.33**
Time of day: morning (38)			
$RB(n_{none})$	1.83	2.09	1.44
$CF_{mean} + IB_{max}(n_{none}, f_{max})$	3.37	**5.80**	3.07
$PB + CF_{mean}(n_{none}, f_{borda})$	**3.70**	5.58	**3.17**
Time of day: afternoon (234)			
$RB(n_{none})$	1.43	1.27	1.01
$PB(n_{none})$	**6.04**	**6.04**	**4.23**
Time of day: night (90)			
$RB(n_{none})$	1.64	1.47	1.16
$CF_{mean} + IB_{max}(n_{none}, f_{max})$	5.25	**5.76**	**4.08**
$PB + CF_{mean} + LB_{max}(n_{sumto1}, f_{mean})$	**5.61**	5.48	4.06

V. Conclusions and Outlook

We conducted first experiments to analyze various recommendation approaches and variants for the task of recommending music to fans of classical music. We identified best-performing variants in terms of precision, recall, and F-measure when further categorizing listeners with respect to age, country, and temporal listening preference. As part of future work, we will investigate a richer set of user characteristics and look deeper into the influence of the investigated normalization and fusion techniques. We will also investigate if results generalize to listeners of other music genres.

Acknowledgments

This research is supported by the EU-FP7 project no. 601166 ("PHENICX") and by the Austrian Science Fund (FWF): P25655.

References

[1] G. Adomavicius and A. Tuzhilin. *Recommender Systems Handbook*, chapter Context-Aware Recommender Systems, pages 217–253. Springer, 2011.

[2] L. Baltrunas, M. Kaminskas, B. Ludwig, O. Moling, F. Ricci, K.-H. Lüke, and R. Schwaiger. InCarMusic: Context-Aware Music Recommendations in a Car. In *Proc. EC-Web*, 2011.

[3] O. Celma. *Music Recommendation and Discovery – The Long Tail, Long Fail, and Long Play in the Digital Music Space*. Springer, 2010.

[4] Z. Cheng and J. Shen. Just-for-Me: An Adaptive Personalization System for Location-Aware Social Music Recommendation. In *Proc. ICMR*, 2014.

[5] J.-C. de Borda. Mémoire sur les élections au scrutin. *Histoire de l'Académie Royale des Sciences*, 1781.

[6] K. Farrahi, M. Schedl, A. Vall, D. Hauger, and M. Tkalčič. Impact of Listening Behavior on Music Recommendation. In *Proc. ISMIR*, 2014.

[7] M. Kaminskas, F. Ricci, and M. Schedl. Location-aware Music Recommendation Using Auto-Tagging and Hybrid Matching. In *Proc. RecSys*, 2013.

[8] C. Liem, M. Müller, D. Eck, G. Tzanetakis, and A. Hanjalic. The Need for Music Information Retrieval with User-centered and Multimodal Strategies. In *Proc. MIRUM*, 2011.

[9] M. Schedl, E. Gómez, and J. Urbano. Music information retrieval: Recent developments and applications. *Foundations and Trends in Information Retrieval*, 8(2–3):127–261, 2014.

[10] M. Schedl, D. Hauger, K. Farrahi, and M. Tkalčič. On the Influence of User Characteristics on Music Recommendation. In *Proc. ECIR*, Vienna, Austria, 2015.

[11] M. Schedl and M. Tkalčič. Genre-based Analysis of Social Media Data on Music Listening Behavior. In *Proc. ISMM*, 2014.

[12] M. Schedl, A. Vall, and K. Farrahi. User Geospatial Context for Music Recommendation in Microblogs. In *Proc. SIGIR*, 2014.

[13] Y. Shi, M. Larson, and A. Hanjalic. Collaborative Filtering Beyond the User-Item Matrix: A Survey of the State of the Art and Future Challenges. *ACM Comput. Surv.*, 47(1):3:1–3:45, May 2014.

[14] X. Wang, D. Rosenblum, and Y. Wang. Context-aware mobile music recommendation for daily activities. In *Proc. ACM Multimedia*, 2012.