Distance- and Rank-based Music Mainstreaminess Measurement

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

Christine Bauer
Johannes Kepler University Linz
Department of Computational Perception
Linz, Austria
christine.bauer@jku.at

ABSTRACT

A music listener's *mainstreaminess* indicates the extent to which her listening preferences correspond to those of the population at large. However, formal definitions to quantify the level of mainstreaminess of a listener are rare and those available define mainstreaminess based on fractions between some kind of individual and global listening profiles. We argue, in contrast, that measures based on a modified version of the well-established *Kullback-Leibler (KL) divergence* as well as *rank-order correlation coefficient* may be better suited to capture the mainstreaminess of listeners. We therefore propose two measures adopting KL divergence and rank-order correlation and show, on a real-world dataset of over one billion user-generated listening events (LFM-1b), that music recommender systems can notably benefit when grouping users according to their level of mainstreaminess with respect to these two measures. This particularly holds for the frequently neglected listener group which is characterized by low mainstreaminess.

KEYWORDS

music recommender systems; mainstreaminess; user modeling

ACM Reference format:
Markus Schedl and Christine Bauer. 2016. Distance- and Rank-based Music Mainstreaminess Measurement. In *Proceedings of UMAP'17 Adjunct, July 09-12, 2017, Bratislava, Slovakia, ,* 4 pages.
DOI: 10.1145/3099023.3099098

1 INTRODUCTION

Using online music platforms such as YouTube, Spotify, or iTunes, music has become easier to access than ever. Still, this opportunity to access a large number of musical works requires novel mechanisms to support users in choosing from the myriad of available musical works and recordings [22]. Music recommender systems have thus become a significant topic, both in research and industry [7, 23].

Various automatic approaches to music recommendation have been proposed [25]. Thereby, "[t]he success of a music recommender system (RS) depends on its ability to propose the right music, to the right user, at the right moment" [16]. Aiming to understand and model users and to provide them with music recommendations tailored to the respective individual (i.e., personalized music recommendations), manifold factors have been investigated, including demographics [17], user activity [9, 28], listening habits [24], and listening venue [10].

The user feature we focus on in this paper is the recently introduced *music mainstreaminess of a user* [13, 24, 27]. This feature harnesses that music listeners may be characterized in terms of the degree to which their music listening preferences correspond to the ones of the overall population. In other words, the music mainstreaminess of a user describes to which degree a user prefers music items that are currently popular rather than ignoring such popularity trends [24].

While we define mainstreaminess on the level of users, the concept is strongly related to the popularity of artists. As a matter of fact, users ranking high on mainstreaminess listen to a lot of popular artists and vice versa. However, even though popularity-based approaches are widely adopted in the field of music recommender systems (e.g., [9, 15, 29]), harnessing user mainstreaminess is a rather new target of research. Furthermore, formal definitions that quantify mainstreaminess are scarce (e.g., [24, 27]). Existing definitions measure mainstreaminess based on fractions between some kind of individual and global listening profiles.

However, we argue that there are better ways to capture a user's mainstreaminess since fraction-based approaches do not take into account the so-called "superstar" phenomenon (also known as "long-tail" or "'hit-driven" phenomenon), which is evident in particular in online music platforms. This phenomenon describes that relatively small numbers of items (the head) dominate the market, while there is a considerable long tail of less popular items [3, 6, 7, 20]. This yield to a disproportionately higher influence of absolute top hits (the head) in fraction-based definitions of mainstreaminess.

Calling on this, we propose and evaluate two novel user mainstreaminess measures that may be better suited to capture a listener's mainstreaminess beyond the very top items. We argue that approaches based on *rank-order correlation* and *Kullback-Leibler (KL) divergence* do *not* overly privilege the very top items since the former considers solely the rank of the items rather than their absolute or relative popularity. The later considers the logarithm of the quotient between individual and global popularity, thereby also penalizing exorbitant disparities between the two. Analyzing the performance of the two proposed mainstreaminess measures on the publicly available LFM-1b dataset [22] shows that personalized music recommendation can notably benefit when grouping users according to their level of mainstreaminess with respect to the proposed two measures.

The remainder of the paper is organized as follows. In Section 2, we briefly review existing literature on music mainstreaminess. We then detail our proposed measures in Section 3. Section 4 shows

UMAP'17 Adjunct, July 09-12, 2017, Bratislava, Slovakia

ACM ISBN 978-1-4503-5067-9/17/07...$15.00
http://dx.doi.org/10.1145/3099023.3099098

how to exploit the measures in collaborative filtering recommendation, discusses results, and provides a comparison to other work. Eventually, we round off the paper in Section 5 with a conclusion and directions for future research.

2 RELATED WORK

Literature in the field of popular music studies and popular music cultures frequently resort to the term *mainstream* (cf. [4]). Often, though, the mainstream is referred to with other terms and phrases (e.g., *hits* [7] or *the head* [12]) to circumscribe the phenomenon, for instance, the hit-driven paradigm [7], the long-tail concept [7, 8], etc. Essentially, all these circumscriptions have in common that they reference to the fact that there is a high concentration of playcounts on the most popular music items (the head), while there exists at the same time a long tail of less popular items (cf. [6, 7]).

In the context of music recommender systems research, the listener-centric feature of user mainstreaminess is a rather new target of research [13, 24, 27]. User mainstreaminess is thereby used to analyze a listener's preferences of music items and compare it with the overall preferences. Other models to describe a listener's music consumption behavior for providing music recommendations include features such as serendipity [31], novelty [11], familiarity [5], unexpectedness [1], or listening intention [5].

Exploiting the mainstreaminess feature in the recommendation process is related to popularity-based recommendation. Such popularity-based recommender systems are widely adopted in numerous domains, including music [9, 15, 29], news [30], or product recommendation in e-commerce in general [2].

Closest to the paper at hand are the works presented in [27] and [24], which both propose formal measures capturing a user's mainstreaminess (spelled "mainstreamness" in [27]) and analyze the recommendation performance of these, among other features. Our work significantly differs from previous works as we counteract the mentioned disproportionate privileging of top music items by proposing a distance- and a rank-based music mainstreaminess measure, which is detailed in the following section.

3 MAINSTREAMINESS DEFINITIONS

The proposed mainstreaminess measures are defined on *preference profiles*, which we compute on a global scale, i.e. considering the entire population of listeners, and on an individual scale, confined to the target user u. We first define the artist frequency $AF_{a,u}$ as the sum of listening events to tracks by artist a listened to by user u. Accordingly, we define AF_a as the total number of listening events to tracks by artist a listened to by the entire population in the dataset under consideration.[1]

Computing the artist frequencies for all artists listened to results in a high-dimensional feature vector, in which each dimension corresponds to the frequency of a particular artist. We refer to this representation of a user's or the global artist frequencies as *preference profile*. Given the LFM-1b dataset [22], which we use in our experiments, these profiles are 585,095-dimensional vectors over all artists in the dataset.

[1]This definition of *artist frequency* corresponds to that of *playcount of an artist*, which is occasionally used in other works.

Artist	Artist Frequency
The Beatles	2,985,509
Radiohead	2,579,453
Pink Floyd	2,351,436
Metallica	1,970,569
Muse	1,896,941
Arctic Monkeys	1,803,975
Daft Punk	1,787,739
Coldplay	1,755,333
Linkin Park	1,691,122
Red Hot Chili Peppers	1,627,851

Table 1: Artists with highest frequency in the dataset.

Exploiting the preference profiles, we propose two mainstreaminess measures for a user u's music taste: symmetrized Kullback-Leibler (KL) divergence (D_u) and rank-order correlation according to Kendall's τ (R_u). KL divergence is a well-established method to compare distributions, which are discrete preference profiles in our case. The use of rank correlation is motivated by the fact that converting feature values to ranks has already been proven successful for music similarity tasks [19, 26]. In addition, we investigate a third, fraction-based (F_u), measure as baseline, which we adopted from previous literature [24]. The respective formal definitions are given in Equations 1, 2, and 3, where A is the set of artists in the dataset, $\widehat{AF_a}$ denotes the normalized artist frequency AF_a (sum-to-unity over all artist frequencies), $\widehat{AF_{a,u}}$ defined accordingly; $ranks(PP_u)$ denotes a function that converts the real-valued preference profile (vector over artist frequencies) of user u to ranks, $ranks(PP_g)$ accordingly on the global level, i.e. considering all users. Please note that we invert the results of the fraction-based formulations and the symmetrized KL divergences in order to be consistent in that higher values indicate closer to the mainstream, whereas lower ones indicate farther away from the mainstream.

$$D_u = \frac{1}{\frac{1}{2} \cdot \left(\sum_{a \in A} \widehat{AF_{a,u}} \cdot \log \frac{\widehat{AF_{a,u}}}{\widehat{AF_a}} + \sum_{a \in A} \widehat{AF_a} \cdot \log \frac{\widehat{AF_a}}{\widehat{AF_{a,u}}} \right)} \quad (1)$$

$$R_u = \tau\left(ranks\left(PP_u\right), ranks\left(PP_g\right)\right) \quad (2)$$

$$F_u = 1 - \frac{1}{|A|} \cdot \sum_{a \in A} \frac{|\widehat{AF_{a,u}} - \widehat{AF_a}|}{\max\left(\widehat{AF_{a,u}}, \widehat{AF_a}\right)} \quad (3)$$

4 MUSIC RECOMMENDATION EXPERIMENTS

In line with common recommender systems evaluation, we perform rating prediction experiments. We use the LFM-1b dataset of user-generated listening events from Last.fm [22] to assess the potential of the proposed mainstreaminess measures. In particular, we analyze the performance of a model-based collaborative filtering approach when tailoring the recommendations to user groups defined according to their level of mainstreaminess.

The LFM-1b dataset's user-artist-playcount matrix (UAM) contains listening events of 120,175 users to 585,095 unique artists. This matrix reflects 1,088,161,692 individual listening events, the distribution of which corresponds to a typical long-tail distribution [7].

Group	RMSE	MAE
Complete UAM	29.105	25.202
D_{low}	74.842	69.495
D_{mid}	5.305	1.919
D_{high}	3.310	1.190
D_{avg}	27.819	24.201
R_{low}	28.349	25.183
R_{mid}	4.476	1.520
R_{high}	6.650	2.410
R_{avg}	13.158	9.704
F_{low}	97.456	92.608
F_{mid}	4.860	1.689
F_{high}	4.221	1.415
F_{avg}	35.512	31.904

Table 2: Root mean square error (RMSE) and weighted mean absolute error (MAE) employing SVD on the playcounts scaled to [0, 1000], for various mainstreaminess definitions (*F*: fraction-based, *D*: KL-divergence-based, *R*: rank-based) and levels (*low*: users in the lower tertile, *mid*: users in the mid tertile, *high*: users in the upper tertile). Additionally, for each measure, the average over the three user groups is reported.

Note that the global population is in our case the Last.fm users in the dataset. Table 1 lists the overall top artists in the considered dataset.

4.1 Experimental Setup

In order to perform the rating prediction task, we first normalize and scale the playcount values in the UAM of the LFM-1b dataset to the range [0, 1000] for each user individually, assuming that higher numbers of playcounts indicate higher user preference for an artist. We then apply singular value decomposition (SVD) according to [21], equivalent to probabilistic matrix factorization, to factorize the UAM and in turn effect rating prediction. In 5-fold cross-validation experiments, we use root mean square error (RMSE) and mean absolute error (MAE) as performance measures.

To obtain an overall performance score, independent of mainstreaminess information, we first conduct an experiment using the set of all users (the full UAM) and report results of the error measures in the first row of Table 2. To investigate the influence of the different mainstreaminess *definitions* and mainstreaminess *levels* on recommendation performance, we then create subsets of users for each combination of mainstreaminess measure and level. For this purpose, we split the users into three equally sized subsets according to their mainstreaminess value: *low* corresponds to users in the lower 3-quantile (tertile) w.r.t. the respective mainstreaminess definition, *mid* and *high*, respectively, to the mid and upper tertile.

4.2 Results and Discussion

Table 2 shows the resulting error measures (RMSE and MAE) for different *definitions* and *levels* of mainstreaminess. We concentrate our discussion on RMSE since it is the more common measure and treats larger differences between predicted and true ratings disproportionately more severe than smaller ones.

As a first observation, we see that the results for the upper 67% of users w.r.t. mainstreaminess (i.e. groups *mid* and *high*) are considerably better than those realized on the entire population (first row), irrespective of the mainstreaminess definition. Their RMSE all rank between 3.3 and 6.6 on the [0,1000]-scaled ratings. As a second observation, the error typically decreases with increasing mainstreaminess of users, which does not come as a surprise since it is easier to predict ratings for users listening to globally popular artists, for which the factorization algorithm can hence learn from a larger amount of data.

Comparing the performance of the three mainstreaminess measures, we see the following: while the fraction-based approach (*F*) performs well on mid and high mainstreaminess listeners, it considerably underperforms on the low mainstreaminess group (RMSE of 97.5). We consider this low mainstreaminess group particularly important, though, for two reasons: (i) it is the most challenging group for recommendation algorithms and (ii) taking a business perspective, low mainstreaminess users are often music aficionados with a quite specific music taste and are presumably willing to spend more money on music than the average listener.[2] On this important group, the KL-based measure (*D*) performs slightly better (RMSE of 74.8) than the fraction-based, but still much worse than the best-performing rank-based (*R*) measure in our study (RMSE of 28.3). The rank-based measure also outperforms the overall results obtained on the entire user set, even on low mainstreaminess listeners. Still, the rank-based measure performs worst among all three measures on high mainstreaminess users. This may be explained by a negative impact of discretization of the very top items when converting frequencies to ranks, which in turn pretty much equalizes those top items.

Analyzing the overall performance among all three user groups, rows D_{avg}, R_{avg}, and F_{avg} in Table 2 denote the respective arithmetic means of the error functions over the user groups, for the three measures. We observe that the rank-based measure considerably outperforms the others, with a RMSE of 13.2, compared to 27.8 for the KL-based and 35.5 for the fraction-based approach.

To summarize, we conclude that the proposed rank-based approach performs superior, both averaged over all user sets and for the low and mid mainstreaminess users. The high mainstreaminess users are, in contrast, best served by the KL-divergence-based measure.

4.3 Comparison to the State of the Art

Directly comparing the RMSE achieved by our approach with that reported in [27], which is the work closest to ours, is barely feasible, even though Vigliensoni and Fujinaga also use a similar dataset of listening events crawled from Last.fm. However, the authors quantize playcounts into the range [1,5], rather than the [0,1000] scale we employ. Nevertheless, our results suggest that the performance of our best, rank-based approach delivers a new benchmark in mainstreaminess-aware music recommender systems, with a RMSE

[2]From many personal discussions with "low mainstreaminess listeners", it occurs that they are less eager to use (relatively cheaper) music streaming services, instead are willing to spend much more money on physical media, concerts, etc. than "high mainstreaminess listeners".

of 13.1 on a [0,1000] scale. The best RMSE reported in [27] when considering mainstreamness information for recommendation is approximately 0.98 on the much narrower [1,5] scale (cf. approach *u.m.* in Figure 2 of [27]). Relating the two different scales, this error value of 0.98 on the [1,5] scale would approximately translate to 196.2 on our [0,1000] scale.

Comparing the results realized by the proposed two measures, i.e., symmetrized KL divergence and rank-based correlation, to those reported in [24] was already effected above since Schedl and Hauger's approach is reflected in the fraction-based measure we adopt as baseline.

5 CONCLUSIONS AND FUTURE WORK

We proposed two novel measures to quantify the music mainstreaminess of listeners. Unlike existing fraction-based approaches, we adopt Kullback-Leibler divergence and rank-order correlation coefficient (Kendall's τ) to relate listener-specific and global preference profiles. To assess the performance of the proposed measures, we conducted a rating prediction task, employing probabilistic matrix factorization on the LFM-1b dataset of user-generated listening events from Last.fm [22]. We quantified performance via RMSE and MAE for all mainstreaminess definitions and three mainstreaminess levels of users. Our results indicate that in most settings the rank-based mainstreaminess definition substantially outperforms both the KL-based and the fraction-based measures, the latter being considered as baseline. In particular, the important low mainstreaminess user group is best served with the rank-based measure.

In future work, we will investigate how well our results generalize to other datasets, e.g., the Spotify playlists dataset [18] or the Million Musical Tweets Dataset [14]. We will further devise models that consider mainstreaminess at the country level, instead of globally. Furthermore, since this kind of research calls for a user-centric evaluation, we will devise an evaluation strategy on a representative set of users in a real-world setting.

6 ACKNOWLEDGMENTS

This research is supported by the Austrian Science Fund (FWF): P25655 and V579.

REFERENCES

[1] Panagiotis Adamopoulos and Alexander Tuzhilin. 2014. On Unexpectedness in Recommender Systems: Or How to Better Expect the Unexpected. *ACM Transactions on Intelligent Systems and Technology* 5, 4 (2014), 54:1–54:32. https://doi.org/10.1145/2559952

[2] Hyung Jun Ahn. 2006. Utilizing Popularity Characteristics for Product Recommendation. *International Journal of Electronic Commerce* 11, 2 (2006), 59–80. https://doi.org/10.2753/JEC1086-4415110203

[3] Christan Anderson. 2006. *The long tail: Why the future of business is selling less of more.* Hachette Books, New York, NY, USA.

[4] Sarah Baker, Andy Bennett, and Jodie Taylor (Eds.). 2013. *Redefining mainstream popular music.* Routledge, London, United Kingdom.

[5] Dmitry Bogdanov. 2013. *From music similarity to music recommendation: Computational approaches based in audio features and metadata.* PhD. Universitat Pompeu Fabra, Barcelona, Spain.

[6] Erik Brynjolfsson, Yu (Jeffrey) Hu, and Duncan Simester. 2011. Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales. *Management Science* 57, 8 (2011), 1373–1386.

[7] Òskar Celma. 2010. *Music Recommendation and Discovery.* Springer, Berlin, Heidelberg, Germany.

[8] Òscar Celma and Pedro Cano. 2008. From Hits to Niches?: Or How Popular Artists Can Bias Music Recommendation and Discovery. In *Proceedings of the 2nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition.* ACM, New York, NY, USA, 5:1–5:8.

[9] Zhiyong Cheng and Jialie Shen. 2014. Just-for-Me: An Adaptive Personalization System for Location-Aware Social Music Recommendation. In *Proc. ICMR.* ACM, New York, NY, USA, 185:185–185:192. https://doi.org/10.1145/2578726.2578751

[10] Zhiyong Cheng and Jialie Shen. 2015. VenueMusic: A Venue-Aware Music Recommender System. In *Proc. SIGIR.* ACM, New York, NY, USA, 1029–1030. https://doi.org/10.1145/2766462.2767869

[11] Charles L.A. Clarke, Maheedhar Kolla, Gordon V. Cormack, Olga Vechtomova, Azin Ashkan, Stefan Büttcher, and Ian MacKinnon. 2008. Novelty and Diversity in Information Retrieval Evaluation. In *Proc. SIGIR.* ACM, New York, NY, USA, 659–666. https://doi.org/10.1145/1390334.1390446

[12] Paolo Cremonesi, Franca Garzotto, Roberto Pagano, and Massimo Quadrana. 2014. Recommending Without Short Head. In *WWW'14 Companion.* ACM, New York, NY, USA, 245–246. https://doi.org/10.1145/2567948.2577286

[13] Katayoun Farrahi, Markus Schedl, Andreu Vall, David Hauger, and Marko Tkalčič. 2014. Impact of Listening Behavior on Music Recommendation. In *Proc. ISMIR.* 483–488.

[14] David Hauger, Markus Schedl, Andrej Košir, and Marko Tkalčič. 2013. The Million Musical Tweets Dataset: What Can We Learn From Microblogs. In *Proc. ISMIR.*

[15] Rajeev Kumar, BK Verma, and Shyam Sunder Rastogi. 2014. Social popularity based SVD++ recommender system. *International Journal of Computer Applications* 87, 14 (2014), 33–37.

[16] Audrey Laplante. 2014. Improving music recommender systems: what can we learn from research on music tags?. In *Proc. ISMIR.* 451–456.

[17] Han-Saem Park, Ji-Oh Yoo, and Sung-Bae Cho. 2006. A Context-Aware Music Recommendation System Using Fuzzy Bayesian Networks with Utility Theory. In *Proc. FSKD.* Springer, Berlin, Heidelberg, Germany, 970–979. https://doi.org/10.1007/11881599_121

[18] Martin Pichl, Eva Zangerle, and Günther Specht. 2015. Towards a Context-Aware Music Recommendation Approach: What is Hidden in the Playlist Name?. In *Proc. ICDM Workshops.* IEEE, Piscataway, NJ, USA, 1360–1365. https://doi.org/10.1109/ICDMW.2015.145

[19] Tim Pohle, Markus Schedl, Peter Knees, and Gerhard Widmer. 2006. Automatically Adapting the Structure of Audio Similarity Spaces. In *Proc. LSAS.* 66–75.

[20] Sherwin Rosen. 2004. *Markets and diversity.* Harvard University Press, Cambridge, MA, USA.

[21] Ruslan Salakhutdinov and Andriy Mnih. 2007. Probabilistic Matrix Factorization. In *Proc. NIPS.* Curran Associates Inc., USA, 1257–1264.

[22] Markus Schedl. 2016. The LFM-1b Dataset for Music Retrieval and Recommendation. In *Proc. ICMR.* ACM, New York, NY, USA, 103–110. https://doi.org/10.1145/2911996.2912004

[23] Markus Schedl, Emilia Gómez, and Julián Urbano. 2014. Music Information Retrieval: Recent Developments and Applications. *Foundations and Trends® in Information Retrieval* 8, 2-3 (2014), 127–261. https://doi.org/10.1561/1500000042

[24] Markus Schedl and David Hauger. 2015. Tailoring Music Recommendations to Users by Considering Diversity, Mainstreaminess, and Novelty. In *Proc. SIGIR.* ACM, New York, NY, USA, 947–950. https://doi.org/10.1145/2766462.2767763

[25] Markus Schedl, Peter Knees, Brian McFee, Dmitry Bogdanov, and Marius Kaminskas. 2015. *Music Recommender Systems.* Springer, Boston, MA, USA, 453–492. https://doi.org/10.1007/978-1-4899-7637-6_13

[26] Dominik Schnitzer, Arthur Flexer, Markus Schedl, and Gerhard Widmer. 2012. Local and Global Scaling Reduce Hubs in Space. *Journal of Machine Learning Research* 13 (October 2012), 2871–2902.

[27] Gabriel Vigliensoni and Ichiro Fujinaga. 2016. Automatic Music Recommendation Systems: Do Demographic, Profiling, and Contextual Features Improve Their Performance?. In *Proc. ISMIR.* 94–100.

[28] Xinxi Wang, David Rosenblum, and Ye Wang. 2012. Context-aware Mobile Music Recommendation for Daily Activities. In *Proc. ACM Multimedia.* ACM, New York, NY, USA, 99–108. https://doi.org/10.1145/2393347.2393368

[29] Yan Yan, Tianlong Liu, and Zhenyu Wang. 2015. A Music Recommendation Algorithm Based on Hybrid Collaborative Filtering Technique. In *Proc. SMP*, Xichun Zhang, Maosong Sun, Zhenyu Wang, and Xuanjing Huang (Eds.). Springer, Singapore, 233–240. https://doi.org/10.1007/978-981-10-0080-5_23

[30] JungAe Yang. 2016. Effects of Popularity-Based News Recommendations ("Most-Viewed") on Users' Exposure to Online News. *Media Psychology* 19, 2 (2016), 243–271. https://doi.org/10.1080/15213269.2015.1006333

[31] Yuan Cao Zhang, Diarmuid Ó. Séaghdha, Daniele Quercia, and Tamas Jambor. 2012. Auralist: Introducing Serendipity into Music Recommendation. In *Proc. WSDM.* ACM, New York, NY, USA, 13–22. https://doi.org/10.1145/2124295.2124300