Music Cold-start and Long-tail Recommendation: Bias in Deep Representations

Andres Ferraro
Music Technology Group - Universitat Pompeu Fabra
Barcelona, Spain
andres.ferraro@upf.edu

ABSTRACT

Recent advances in deep learning have yielded new approaches for music recommendation in the long tail. The new approaches are based on data related to the music content (i.e. the audio signal) and context (i.e. other textual information), from which it automatically obtains a representation in a latent space that is used to generate the recommendations. The authors of these new approaches have shown improved accuracies, thus becoming the new state-of-the-art for music recommendation in the long tail.

One of the drawbacks of these methods is that it is not possible to understand how the recommendations are generated and what the different dimensions of the underlying models represent. The goal of this thesis is to evaluate these models to understand how good are the results from the user perspective and how successful the models are to recommend new artists or less-popular music genres and styles (i.e. the long tail). For example, if a model predicts the latent representation from the audio but a given genre is not well represented in the collection, it is not probable that the songs of this genre are going to be recommended.

First, we will focus on defining a measure that could be used to assess how successful a model is recommending new artists or less-popular genres. Then, the state-of-the-art methods will be evaluated offline to understand how they perform under different circumstances and new methods will be proposed. Later, using an online evaluation it will be possible to understand how these recommendations are perceived by the users.

Increasingly, algorithms are responsible for the music that we consume, understanding their behavior is fundamental to make sure they give the opportunity to new artists and music styles. This work will contribute in this direction, making it possible to give better recommendations for the users.

CCS CONCEPTS

• **Information systems → Recommender systems**; *Information extraction*; *Music retrieval*.

KEYWORDS

music recommender systems, deep neural networks, collaborative filtering, content-aware recommendation, ethics, popularity bias

RecSys '19, September 16–20, 2019, Copenhagen, Denmark

ACM ISBN 978-1-4503-6243-6/19/09.
https://doi.org/10.1145/3298689.3347052

ACM Reference Format:
Andres Ferraro. 2019. Music Cold-start and Long-tail Recommendation: Bias in Deep Representations. In *Thirteenth ACM Conference on Recommender Systems (RecSys '19), September 16–20, 2019, Copenhagen, Denmark.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3298689.3347052

1 INTRODUCTION

There are multiple factors that make music recommendation a complex problem. Many different aspects could be taken into account when making a recommendation, some could be related to theoretical aspects of the music, others related to ideological connotations or social aspects. Beyond this, some of these aspects are subjective or are not possible to quantify, they change in time or depend on the context in which we measure it.

This is a common problem when we try to define a measure of similarity in music [28]. One might argue that defining a similarity is an oversimplification [6] and is not possible to manually select all the relevant aspects of the music. As an alternative, collaborative filtering methods generate a recommendation by identifying patterns in what people listen from historical information. Therefore, is not required to define this similarity by hand but there are other problems with these methods: since they do not consider any other information than the interactions between user and items is not possible to generate recommendations for new items (cold-start problem) and they tend to follow the distribution of popularity of the music [18], meaning that the most popular items are more recommended (long-tail recommendation problem) [7].

Some solutions had been proposed for long-tail and cold-start recommendation. By combining the best of these solutions we can create hybrid systems that can reduce the problem but integrating multiple systems can be complex [32]. With the advances in deep learning, new methods are able to automatically learn a representation from the data without the need for manually selecting the features. New methods had been proposed [24, 36] that try to find a common representation of the songs that can be generated from multiple types of information related with the music in a multimodal approach (e.g: from the audio of the songs, text extracted from the web like social networks or Wikipedia, crowdsourced tags related to songs or other metadata).

Still, these solutions present some issues, e.g: related to the fact that they work as black-boxes, for example, it is difficult to explain the results and it is hard to know if the different musical styles are equally represented. Also, previous works do not show 1) how robust these methods are to biased datasets 2) if it is possible to generate recommendations to completely new styles or genres that are less present in the user-item interactions 3) if it is possible with these methods to consider temporal aspects of the music trends, how users change the musical taste or if it is possible to reduce filter bubbles with these methods.

The growth of music streaming services in the last years has increased the importance of music recommender systems, reducing the choice overload is commonly referred to as one of the advantages of these systems. But is important to understand the increasing impact that these systems have in what people listens. They define which song will be the next hit, how much will an artist earn or even music genres that get almost zero promotion with the risk of fall in forgetfulness. This raises some ethical issues that had been discussed in previous works, Holzapfel et al. [15] raise the question if a group of artists that are never recommended by a system can be considered a case of discrimination. As researchers, we have to think about the implications of the systems we develop and the importance of assuring every artist a fair chance to reach the public. In this line of works, lately, some researchers started to propose methods to address these issues [22, 23].

We expect that all artists in a collection can have the opportunity of being listened to, and then users can decide if they want to continue or not. Therefore, in this thesis, we consider the music recommendations as a multistakeholder problem [1] and we propose to evaluate how successful the current systems considered state-of-the-art are recommending new and less-popular genres and styles while at the same time not having a negative impact on the user's experience. New approaches will be explored with the focus of recommending new and underrepresented genres, allowing users to get more diverse recommendations without reducing user satisfaction in the long-term instead of maximizing the instant satisfaction.

The remainder of this paper is organized as follows. In Section 2 a review of the related works is presented, in Section 3 the proposed approach is described. In Section 4 contains a description of the datasets to be used. Finally, Section 5 gives the conclusions.

2 RELATED WORK

2.1 Metrics and Evaluation

The problem of recommending music in the long-tail was first raised by Celma [7] and proposed multiple novel ideas that were followed later by other researchers. The author proposed metrics to evaluate a music recommender system taking into account the popularity of the items [8]. It shows that using a recommender system based on last.fm[1] data the results reinforce popular artists and discard less known music.

There are multiple methods for combining recommender systems with the goal of maximizing a given metric, which could lead to increasing the accuracy and diversity at the same time [3, 4, 10].

In offline evaluation, there are multiple metrics that had been proposed to evaluate recommender systems. For measuring the accuracy of the system, some metrics are Root Mean Squared Error or Mean Average Error [27, 32]. These metrics could be used when we want the system to predict the exact same value rated by the user. This could be the case of explicit ratings, or if we transform from the implicit data to a scale that we want to predict. Other systems instead of predicting the exact value, focus on the ranking between

[1]https://www.last.fm

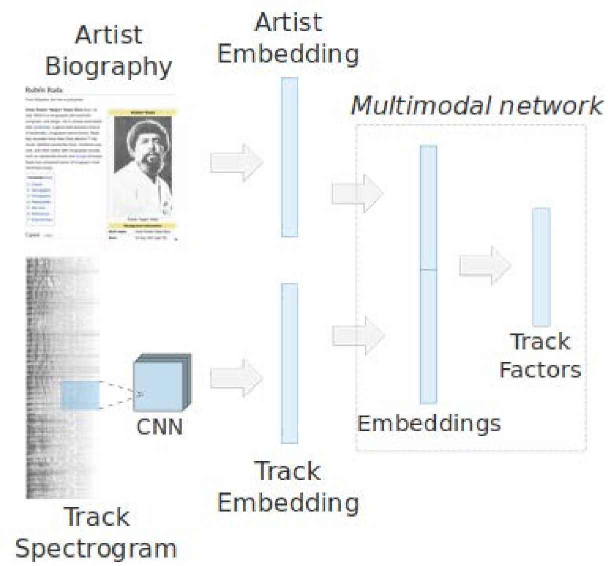

Figure 1: Multimodal architecture for cold-start music recommendation [24].

the items since the user perceives better accurate prediction on the highly rated items rather than on the lower rated predictions. For this purpose some metrics from the field of Information Retrieval are used, for example, Mean Average Precision, Normalized Discounted Cumulative Gain, Mean Reciprocal Rank or Precision and Recall at some cut-off [32].

Other works show that novelty and diversity are very important to give good recommendations and increase user satisfaction [40]. In the same line, Knijnenbur et al. [20] argue that measuring the algorithmic accuracy is insufficient to analyze the user experience and propose a framework that can be used as a guideline for the evaluation of the recommendations from the user perspective. This framework considers multiple aspects from the algorithms and the recommendations (objective and subjective) that define how the user perceives them and makes it possible to analyze the user experience.

Finally, there are also other metrics used for offline evaluation to compare multiple recommender systems related to the diversity or serendipity of the results [2, 16, 29, 32, 37, 38].

2.2 Methods for Recommendation

Given the complexity of the music domain, it is clear that to give good recommendations in the long-tail or cold-start is important to have multiple sources of data or formats [24] that can cover as many aspects as possible from the music. Then, the problem is how to extract relevant information that can be used for the recommendations and how to combine it.

In order to make recommendations, multiple studies try to define a similarity measure [11, 18, 19] from different aspects related to the music.

Some recent methods automatically learn a representation from the content or context of the music. The recent advances in this direction showed good accuracy of audio-based recommendations [36], text and semantic information for the recommendations [24] (shown in Figure 1).

In the task of playlist generation and continuation, many new approaches had been proposed for the RecSys Challenge [9]. In particular, the 'creative' track of the challenge was restricted for solutions that other sources of information apart from the data provided by the organizers. Therefore, some researchers proposed solutions using multimodal approaches [12, 39] by learning embeddings for the playlists, artist or songs.

Sequence-aware recommender systems [26] also could be applied to music recommendation, and some methods are based on learning embeddings or latent representations for the items, especially the ones based on recent advances in deep learning. In this field, Vall et al. [34] study gives an example of using different types of information for improving the playlist generation from songs better represented, especially for very infrequent songs. Also, Vall et al. [35] investigate the importance of order, song context and popularity bias in music playlist task.

It is also relevant to mention some methods that incorporate information from the user context [13, 30, 31] to give better recommendations, this could allow us to understand the user needs at a particular moment, in this way we could reduce the effect of filter bubbles and it proves the importance of considering the temporal aspect of the recommendation [28].

3 PROPOSED APPROACH

The proposed approach is divided into two parts, one focused on the metrics and the other more related to the methods for generating the recommendations. In summary, these are the main task that will be performed, which are explained in detail in the following subsections:

- Compare and define metrics to assess if a model recommends new and less-popular music styles and genres.
- Compare current state-of-the-art methods for cold-start and long-tail recommendations according to these metrics.
- Obtain better representations of the tracks to improve the results.

These tasks are initially carried out doing an offline evaluation with a grand-truth. Since the final goal is to compare the performance of these metrics with how the users perceives the recommendations, after these tasks, an online test will be carried out.

3.1 Metrics

With the goal of having a better understanding of the systems behavior in different situations, the first task is to define an overall metric that can assess how probable is for new or less popular artists to be recommended and measure how it changes across the different genres. Is also important to know how many different users each artist is recommended to. Otherwise, all artists could be recommended to one single user which is not much better.

For this purpose, the Gini coefficient could be used to measure how the genres and artists recommendations are distributed across the users, but it is possible that a single measure is not enough to completely describe the behavior. If we consider the system as a multistakeholder recommender, there are also some metrics that could be used to understand the behavior of the according to the providers of the content [1]. In our case, this could be applied to the artists, e.g: the exposure of the artists across the different genres or the ratio of hits for the genres.

At the same time, other known measures like accuracy, diversity and novelty of the results should be measured to see how they are related. These measures will allow at the end to see the impact that they have on the user's experience and to understand how the users perceive the different recommender systems.

3.2 Baseline Methods Evaluations

After defining the measures that will be used to know if the recommendations are better for the new and less-known artists, the next task is to evaluate the current state-of-the-art methods for cold-start and long-tail recommendations according to the accuracy. In particular, the method proposed by Van den Oord et al. [36] based on audio and the method proposed by Oramas et al. [24] based on text and metadata. The authors of these works show that the methods are more accurate, but the goal of this thesis is to evaluate how much popular tracks or genres influence the representations and therefore the final recommendations. In concrete, is it possible with these methods to recommend a completely new genre? or with what degree of success, it recommends genres that are in the long tail?

In order to evaluate the methods offline, after reproducing these approaches, the idea is to manually reduce the presence of some genres or artists in the training set and then evaluate how successful the systems are on recommending these items.

3.3 Propose Additional Methods

The next challenge is to propose a way to give recommendations that are better distributed across the different genres and musical styles, that is, more robust to the popularity and bias from the dataset, while at the same time not negatively affecting the user's experience. For this, we will explore different ways of obtaining better representations of the tracks that could allow us to maximize the desired metrics.

In previous works, different sources of information as text, image, and audio [24, 36] are combined to obtain representations of the tracks that achieve a higher accuracy of the recommendations. After evaluating these methods for the desired metrics we will see if combining multiple sources also give better performance according to the desired metrics. If this is the case, new sources of information that capture other aspects of the music could be incorporated to improve the representations further, for example:

- Following the approach proposed by Hamilton et al. [14] embeddings of the artists could be obtained from a graph of relations (e.g: from artist collaborations) and combined in the representation used by Oramas et al. [24].
- Different representations from the lyrics of the songs could be obtained using Bag-of-words, word2vec or with more recent methods of documents representation [17].

An alternative way of generating better representations of the tracks could be by putting restrictions to the latent space that could allow the algorithms to generalize better to unseen genres or with fewer tracks. For example, previous works on Prototypical Networks [33] give good results on classifying with few elements [25] (i.e: few-shot and zero-shot learning). We propose to use this approach to obtain a representation from the audio that could allow us to generate recommendation on musical styles or genres with few examples. This network learns a prototypical representation of the classes and locates the elements closer to their class. Therefore, another advantage of this approach is that gives better interpretability of the model, which can be used to explain the recommendations to the users.

In a different line of solutions, we propose to consider this problem from the multistakeholder perspective [1], in particular, as a multi-objective optimization problem, where we define the metrics that we want to optimize for each stakeholder: the users and the artists. In our case, could be applied in the following ways:

- Re-ranking of the recommendations of a system (that was trained to maximize the accuracy of the results) to increase the artists' exposure
- Restricting the system to generate recommendations that are independent of the popularity of the items, Abdollahpouri et al. [1] summarize 3 approaches for this under the title of "Fairness from independence".

3.4 User-centric evaluation

The offline evaluation will allow us to analyze the distribution of the recommendations and the accuracy, but since we have a ground truth biased to the most popular music, we propose to carry on an online test following the framework proposed by Knijnenbur et al. [20]. This test will allow us to have a better idea about the users experience and how they perceive the different recommender systems.

For the experiment, we will first request to the users to enter a list of tracks, that will be used as input for each system to generate the recommendations. These recommendations will be presented to the users independently, they will be asked to listen to the tracks and give a rating. Also, each user will be requested to answer a questionnaire. From the answers, we will model the users' behavior and analyze how they perceive the systems.

From the results, we will compare the influence of changing the distribution of the recommendations in the satisfaction reported by the users and how this is related to some personal and contextual aspects.

3.5 Additional Tasks

Playlists are one of the ways that users currently consume more music, therefore is also important to know the performance of the current methods for playlist generation and continuation for new artists in the different genres.

After comparing the different metrics we could evaluate current methods for playlists generation to know if the degree in which they include long-tail artists and genres.

4 DATASETS

The Million Song Dataset (MSD) [21] is a large dataset of audio features with metadata, it was expanded by the Music Information Retrieval community with additional information including tags, lyrics and other annotations. The Echo Nest Taste Profile Subset [5] is related to the MSD, providing play count of 1 million users and 380,000 songs. Also, at some point, it was possible to download the audio previews for the MSD from 7digital.com, these previews consist of audio samples between 7 and 30 seconds of the songs. The audio fragments vary in their quality, encoded as MP3 with a bitrate ranging from 64 to 128 Kbps and the sample rates of 22 KHz or 44 KHz. Additionally, the Last.fm Dataset [5] is related to MSD and provides tags in the song level, which are extracted from last.fm. The tags were crowdsourced and cover genre, instrumentation, moods and eras.

The MSD-A dataset [24] is also related to MSD and contains biographies of the artists, this dataset was collected and used to obtain the representations of the artists for cold-start recommendation. The dataset consists of 328,821 tracks from 24,043 artists with a biography of at least 50 characters long. Each artist has at least one tag associated.

5 CONCLUSIONS

The goal of this research is to contribute to the understanding of some models that are currently proposed as state-of-the-art for cold-start and long-tail music recommendation. These models combine different types of data, e.g: audio and user listening history, to obtain latent representations of the items that are used to generate the recommendations. Evaluating the accuracy of the models only show that they can predict the recommendations in the ground-truth with some success, but this ground-truth also is biased following the long-tail distribution, therefore is necessary to carry a deeper evaluation that could give a better understanding of the recommendations from the users perspective.

By analyzing how these models perform for different music styles or genres and in different levels of popularly we will understand better the performance they have for recommending the long-tail.

After having a better understanding of the current solutions, new systems will be proposed exploring different approaches. For example, including more information in the models to capture different aspects of the music. Another approach is to use other methods that, by adding some restrictions to the model, could help to give better-distributed recommendations. A third approach based multi-objective optimization will try to maximize at the same time a metric for the users and another metric for the artists.

Finally, a user-centric evaluation will be carried on to compare the performance of the recommender systems from the users perspective. This will allow us to better analyze the influence of the different approaches in the experience of the users taking into consideration some personal and contextual aspects.

ACKNOWLEDGMENTS

This work is conducted under the supervision of Prof. Xavier Serra. Special thank to Frederic Font, Sergio Oramas and Dmitry Bogdanov for their valuable feedback. The study is partially supported by Kakao Corp.

REFERENCES

[1] Himan Abdollahpouri, Gediminas Adomavicius, Robin Burke, Ido Guy, Dietmar Jannach, Toshihiro Kamishima, Jan Krasnodebski, and Luiz Pizzato. 2019. Beyond Personalization: Research Directions in Multistakeholder Recommendation. *arXiv preprint arXiv:1905.01986* (2019).

[2] Panagiotis Adamopoulos and Alexander Tuzhilin. 2015. On unexpectedness in recommender systems: Or how to better expect the unexpected. *ACM Transactions on Intelligent Systems and Technology (TIST)* 5, 4 (2015), 54.

[3] Alejandro Bellogín, Iván Cantador, and Pablo Castells. 2010. A study of heterogeneity in recommendations for a social music service. In *Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems*. 1–8.

[4] Alejandro Bellogín, Iván Cantador, and Pablo Castells. 2013. A comparative study of heterogeneous item recommendations in social systems. *Information Sciences* 221 (2013), 142–169.

[5] Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman, and Paul Lamere. 2011. The Million Song Dataset. In *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR)*. 591–596.

[6] Donald Byrd and Tim Crawford. 2002. Problems of music information retrieval in the real world. *Information processing & management* 38, 2 (2002), 249–272.

[7] Òscar Celma. 2009. *Music recommendation and discovery in the long tail*. Ph.D. Dissertation. Universitat Pompeu Fabra.

[8] Òscar Celma and Perfecto Herrera. 2008. A new approach to evaluating novel recommendations. In *Proceedings of the 2nd ACM Conference on Recommender Systems (RecSys)*. 179–186.

[9] Ching-Wei Chen, Paul Lamere, Markus Schedl, and Hamed Zamani. 2018. Recsys challenge 2018: automatic music playlist continuation. In *Proceedings of the 12th ACM Conference on Recommender Systems (RecSys)*. 527–528.

[10] Andres Ferraro, Dmitry Bogdanov, Kyumin Choi, and Xavier Serra. 2019. Using offline metrics and user behavior analysis to combine multiple systems for music recommendation. *arXiv preprint arXiv:1901.02296* (2019).

[11] Andres Ferraro, Dmitry Bogdanov, and Xavier Serra. 2019. Skip prediction using boosting trees based on acoustic features of tracks in sessions. *arXiv preprint arXiv:1903.11833* (2019).

[12] Andres Ferraro, Dmitry Bogdanov, Jisang Yoon, KwangSeob Kim, and Xavier Serra. 2018. Automatic playlist continuation using a hybrid recommender system combining features from text and audio. In *Proceedings of the ACM Recommender Systems Challenge 2018*. 2.

[13] Andreas Forsblom, Petteri Nurmi, Pirkka ÃĚman, and Lassi Liikkanen. 2012. Out of the bubble: serendipitous even recommendations at an urban music festival. (2012), 253–256.

[14] William L Hamilton, Rex Ying, and Jure Leskovec. 2017. Representation learning on graphs: Methods and applications. *arXiv preprint arXiv:1709.05584* (2017).

[15] Andre Holzapfel, Bob Sturm, and Mark Coeckelbergh. 2018. Ethical dimensions of music information retrieval technology. *Transactions of the International Society for Music Information Retrieval (TISMIR)* (2018).

[16] Rong Hu and Pearl Pu. 2011. Helping Users Perceive Recommendation Diversity.. In *Proceedings of the Workshop on Novelty and Diversity in Recommender Systems (DiveRS), at the 5th ACM International Conference on Recommender Systems (RecSys)*. 43–50.

[17] Mehran Kamkarhaghighi, Eren Gultepe, and Masoud Makrehchi. 2019. *Deep Learning for Document Representation*. 101–110.

[18] Peter Knees and Markus Schedl. 2013. A survey of music similarity and recommendation from music context data. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 10, 1 (2013), 2.

[19] Peter Knees and Markus Schedl. 2016. Collaborative Music Similarity and Recommendation. In *Music Similarity and Retrieval*. 179–211.

[20] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction (UMUAI)* 22, 4-5 (2012), 441–504.

[21] Brian McFee, Thierry Bertin-Mahieux, Daniel PW Ellis, and Gert RG Lanckriet. 2012. The million song dataset challenge. In *Proceedings of the 21st International Conference on World Wide Web (WWW)*. 909–916.

[22] James McInerney, Benjamin Lacker, Samantha Hansen, Karl Higley, Hugues Bouchard, Alois Gruson, and Rishabh Mehrotra. 2018. Explore, exploit, and explain: personalizing explainable recommendations with bandits. In *Proceedings of the 12th ACM Conference on Recommender Systems (RecSys)*. 31–39.

[23] Rishabh Mehrotra, James McInerney, Hugues Bouchard, Mounia Lalmas, and Fernando Diaz. 2018. Towards a fair marketplace: Counterfactual evaluation of the trade-off between relevance, fairness & satisfaction in recommendation systems. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM)*. 2243–2251.

[24] Sergio Oramas, Oriol Nieto, Mohamed Sordo, and Xavier Serra. 2017. A deep multimodal approach for cold-start music recommendation. In *Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems*. 32–37.

[25] Jordi Pons, Serrá Joan, and Serra Xavier. 2019. Training Neural Audio Classifiers with Few Data. In *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 16–20.

[26] Massimo Quadrana, Paolo Cremonesi, and Dietmar Jannach. 2018. Sequence-aware recommender systems. *ACM Computing Surveys (CSUR)* 51, 4 (2018), 66.

[27] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2015. *Recommender Systems Handbook*. Springer.

[28] Markus Schedl, Emilia Gómez, Julián Urbano, et al. 2014. Music information retrieval: Recent developments and applications. *Foundations and Trends® in Information Retrieval* 8, 2-3 (2014), 127–261.

[29] Markus Schedl and David Hauger. 2015. Tailoring music recommendations to users by considering diversity, mainstreaminess, and novelty. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 947–950.

[30] Markus Schedl, David Hauger, and Dominik Schnitzer. 2012. A model for serendipitous music retrieval. In *Proceedings of the 2nd Workshop on Context-awareness in Retrieval and Recommendation (CaRR)*. 10–13.

[31] Markus Schedl and Dominik Schnitzer. 2013. Hybrid retrieval approaches to geospatial music recommendation. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. 793–796.

[32] Markus Schedl, Hamed Zamani, Ching-Wei Chen, Yashar Deldjoo, and Mehdi Elahi. 2018. Current challenges and visions in music recommender systems research. *International Journal of Multimedia Information Retrieval (IJMIR)* 7, 2 (2018), 95–116.

[33] Jake Snell, Kevin Swersky, and Richard Zemel. 2017. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*. 4077–4087.

[34] Andreu Vall, Matthias Dorfer, Hamid Eghbal-zadeh, Markus Schedl, Keki Burjorjee, and Gerhard Widmer. 2019. Feature-combination hybrid recommender systems for automated music playlist continuation. *User Modeling and User-Adapted Interaction* (2019).

[35] Andreu Vall, Massimo Quadrana, Markus Schedl, and Gerhard Widmer. 2019. Order, context and popularity bias in next-song recommendations. *International Journal of Multimedia Information Retrieval (IJMIR)* 8, 2 (2019), 101–113.

[36] Aaron Van den Oord, Sander Dieleman, and Benjamin Schrauwen. 2013. Deep content-based music recommendation. In *Advances in neural information processing systems*. 2643–2651.

[37] Saul Vargas and Pablo Castells. 2011. Rank and relevance in novelty and diversity metrics for recommender systems. In *Proceedings of the 5th ACM Conference on Recommender Systems (RecSys)*. 109–116.

[38] Saúl Vargas and Pablo Castells. 2013. Exploiting the diversity of user preferences for recommendation. In *Proceedings of the 10th conference on open research areas in information retrieval (OAIR)*. 129–136.

[39] Hamed Zamani, Markus Schedl, Paul Lamere, and Ching-Wei Chen. 2018. An Analysis of Approaches Taken in the ACM RecSys Challenge 2018 for Automatic Music Playlist Continuation. *arXiv preprint arXiv:1810.01520* (2018).

[40] Cai-Nicolas Ziegler, Sean M McNee, Joseph A Konstan, and Georg Lausen. 2005. Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web (WWW)*. 22–32.