State-Of-The-Art Music Recommendation

Katharina Stengg

University of Klagenfurt
`katharina.stengg@aau.at`

Abstract. Nowadays tons of music is available to almost everyone almost all the time. This highlights the need of a good song recommendation technique to prevent user's from information overload. There exists a variety of methods for this task, some only from a computer scientists' view but also a lot of interdisciplinary work, especially when it comes to include sociological or psychological factors. This work aims to present some of the most recent methods for music recommendation using either traditional or more specific approaches. Another core aspect of this seminar paper is the critical discussion and evaluation with the goal to answer the main hypothesis on whether we already achieve sufficient results using state-of-the-art methods.

Key words: recommender systems, music recommendation, artificial intelligenece, deep learning, lstm, content-based, collaborative, knowledge-based, hybrid

1 Introduction

According to Jannach, Zanker *et al.*[10] the overall goal of recommender systems is to build systems that help users in making decisions.
Recommendations should be applicable to the broad mass of users but also suggest on such a personal level that it outperfoms field experts.
Recommendation systems aim to relieve users from the information overload they need to handle when making a decision without the actual need to consult an expert or even pay for the service.

Im general we differ between the following main categories of recommendation systems:

- Collaborative Filtering
- Content-based Filtering
- Kowledge-based Recommendation
- Hybrid Recommendation

Collaborative Filtering Collaborative Filtering (CF) is based on other user's opinions and ratings.

This recommendation method is similar to people's behaviour in real world. If we like the camera we bought, the friends we tell about might also consider to buy it. If we don't like it, they probably also won't buy it. Collaborative filtering is mainly used to recommend an item, predict the rating for an item or provide contrained suggestions. [11]

Content-based Filtering In comparison to collaborative filtering this approach doesn't rely on user's ratings but on items with similar ratings.
A user profile containing the user's interest is used. [12]

Kowledge-based Recommendation Recommendations are made based on available recommendations and user's demands. [12]

Hybrid Recommendation Each of the methods mentioned above comes with a variety of advantages and disadvantages.
Hybrid recommendation systems are combinations of other recommendation approaches to eliminate certain limitations and improve the overall performance. There is a variety of possibilities to combine methods. Switching between recommendations, combining the scores to an overall score, having various recommendations as output and more. [12]

The main hypothesis of this work is:

State-of-the-art music recommendation: Are we there yet?

This question indicates that there is a lot of new development in the field of music recommendation but we don't have an answer to the really important question of whether we already achieve sufficient results in real-world usage scenarios such as when streaming music on a telephone.

To answer this question this work aims to analyze various recent research approaches for the task of music recommendation.
Nowadays various disciplines work on this topic such as *Computer Science*, *Human Computer Interaction* (HCI), *Psychology*, *Sociology*, *Information Processing* and more. Therefore we can say that the core topic of recommendation systems is broadly spread among various fields and therefore a lot of different aspects need to be taken into account to get high qualitative recommendations.

An additional difficulty of this domain is the human factor. It can be really hard to generalize because of different users and their personality, mood and other factors. Recent work did a lot of research into this direction.
To summarize the following chapters aim to provide an overview of recent developments for the task of music recommendation followed by a critical discussion and evaluation as well as conclusion.

2 State-of-the-art

In 2020 Melchiorre, Zangerle *et al.* proposed their work were they analyzed the personality bias of common algorithms in the area of music recommendation. They stated that other work already proved that there is for example a bias on item or user level in multimedia recommendations. Addtionally there exists work proving a correlation between music preferences and personality traits.
For their experiment they created a novel dataset combining user's personality traits and music consumption. To gather this data they used Twitter microblog posts with certain hashtags as well as the music metadata dataset MusicBrainz. Using the *IBM Personality Insight API* the authors were able to integrate the user's personality traits according to the *OCEAN* model of psychology. The authors compared *Recall@K* and *NDCG@K* (normalized discounted cummulative gain), with K being the length of the recommendation list, of three popular music recommenders namely *SLIM*, *EASE* and *Mult-VAE*. They were able to prove that low scores on personality traits such as *openness*, *extraversion* and *conscientiousness* lead to higher performance while low scores for *neuroticism* and *agreeableness* lead to low performance according to the chosen metrics. Finally they stated some limitations of their work as well as planned extensions. [4]

Ferraro, Serra *et al.* published a work where they analyzed the gender imbalance in song recommendation systems. A core driver of their research were interviews with musicians where the musicians wanted female artists to be recommended more frequently than it currently happens. To analyze whether there is actually a gender bias, the authors incorporated gender information into a dataset and performed recommendations using a collaborative-filtering method named *ALS*. They were able to find that the ranking of results includes a bias where the tendency first shows music from male artists. To improve the situation they suggested a re-ranking method. In the future they want to test their re-ranking method, which is basically a feedback loop, on datasets more similar to real-world usage. [13]

CAME, content and context-aware music embedding for recommendation, is a novel approach proposed by Wang, Zhang *et al.*. They stated that traditional methods often have limitations due to missing additional information aspects. Therefore they aim to comnbine content and contextual information for a personaliezes music recommendation system. Content information is for example metadata, tags or lyrics of songs. Contextual information consists of the user's behaviour such as session when the user listens to tracks or sequences of music played. The combination of those two information types is done via a heterogeneous informatin network (HIN). Their approach *CAME* then uses deep learning techniques such as CNN and attention mechanisms. Evaluation was done qualitativly and quantitativly on a real-world dataset crawled from Xiami music[1]. The authors concluded that their approach outperforms the baseline approaches.

---

[1] https://www.mi.com/global/

CAME differs from other state-of-the-art recommendation systems because they also incorporate dynamic music features, heterogeneous information and various characteristics of music. [1]

Moscato, Picariello *et al.* proposed a novel approach where they included user's moods, emotions and personality traits for the task of music recommendation. According to the authors recent publications revealed evidence for the importance of such aspects in terms of a good recommendation. Their approach extends a conent-based recommendation method with features of user's mood and personality.
The authors used a dataset from Dezzer[2] as well as two sources for user's personality informations. They were able to show that the newly introduced factors are of great importance for good music recommendations. [7]

In 2020 Hansen, Hansen *et al.* tried to tackle challenges such as dealing with short songs or user's listening to the same songs multiple times. For this they used a dataset from Spotify[3] and a neural network architecture.
Their main hypothesis was that contextual information such as the device or a timestamp as well as sequential user behaviour are important dimensions for music recommendation systems. They were able to conclude that both contextual and sequential data are of high value for the core task of music recommendation. [8]

In their work Niyazov, Mikhailova *et al.* proposed a novel approach for content-based music recommendation. They want to recommend suitable music to users based on the acoustic similarity of songs. This means that they use the acoustig signal of a song transformed into a suitable representation. To find the best approach they performed two seperate experiments and afterwards compared the quantitative results.
The first experiment is based on acoustic feature analysis. They extract acoustic features from songs, transform them into a vector representation and use 10 Nearest Neighbors with Euclidean distance to get the 10 best recommendations per song.
The second experiment is using artificial neural networks. The networks are trained using spectogram representations of the acoustic signals. Using the model a vector representation for each song is predicted. Similar to the first approach using 10-NN the best 10 recommendations per song are gathered.
Both approaches got compared with random recommendations using precision, recall, F1-score and a version of the F1-score with a threshold of 0.4. This means that all recommendations below this value are considered as irrelevant. The value was introduced by the authors based on their evaluation.
Comparing the quantitative measurements the authors were able to conclude that both approaches outperformed the random recommendation whit the sec-

---

[2] https://www.deezer.com/de/offers
[3] https://open.spotify.com/intl-de

ond approach using Neural Networks achieving the best results.
Addtionally the authors invesitaged genre-specific clustering via visualizing vector spaces on a plane. Due to genre-specific clustering being closely related to the acoustic sound of songs this again highlighted their findings that approaches using ANN achieve better results. [3]

La Gatta, Moscato *et al.* proposed a novel approach for music recommendation using hypergraph embedding. They split their work into three main parts. First they invesitaged how to model and store the relevant data in a hypergraph-based structure. Afterwards they focussed on generating hypergraph embeddings from their dataset. The third part of their work was the actual recommendation task where they used graph-based machine learning to generate the top-k track recommendations for each user.
One core challenge of their work was the modelling of scenarios such as a song being part of an album, users listening to songs, et cetera. This was done using a hypergraph-based structure which has vertices and edges but also metadata information.
For evaluation La Gatta, Moscato *et al.* compared their novel approach with other state-of-the-art music recommenders such as *CAME* [1]. They were able to surpass them at some quality metrics. [2]
The authors of [6] proposed a novel approach called *DNBMR* (Deep neural based music recommendation) to prevent the cold start problem many other systems suffer from. They incorporated music attributes as well as music temporally liked by user's. With a combination of *LSTM* (long short term memory) and neural networks their approach beat two other baseline recommendation systems. Evaluation was done using a dataset from a large Asian streaming platform and the metrics *Accuracy*, *Recall*, *F-score*, *AUC* (area under curve), *MAP* (mean average precision) and *user coverage.*

In 2023 a novel approach for music recommendation based on user's emotion was published. The authors tried to predict the user's emotion based on the facial expression via a neural network and a webcam livestream. Depending on the emotion a suitable *Spotify*[4] playlist was displayed. For evaluation they used *Accuracy*, *Loss*, *F1-score*, *Precision* and *Recall.*[5]

3   Critical discussion and evaluation

This section aims to provide a critical view onto current methods in the field of music recommendation.

One core aspect is the usage of incomplete data. Not every novel approach uses a suitable and current dataset. Some even pick a small percentage of the data which doesn't model real-world scenarios where we have tons of data and users. Another aspect is the access to data.

---

[4] https://open.spotify.com/intl-de

For example in one work [4] the authors crawled Twitter data to gain insights on the personality and the music usage of users. One of their limitations is that users only post what they want others to know. There probably is a huge bias in that data.

A large percentage of work in the field of recommender systems doesn't perform proper evaluation methods. Approaches such as [3] are considered decent just through analyzing various performance metrics on a test dataset but not through real usage tests. This leads to the problem that many research approaches in the field of recommender systems are not applicable for real-world scenarios. In their work Cremonesi and Jannach evaluated recent work in the field of recommender systems for their reproducibility, evaluation methodologies, baselines and research questions to answer the questions whether they actually improve the state-of-the-art. If they found evidence that there is no real progress, they tried to find reasons for that and provided recommendations for researchers but also reviewers and chairs. They suggested to move from a strong reliance on accuracy and well-explored problems to more practice related work. [9]

Summarizing we can say that every approach has it's own limitations. However this should not prevent researchers from using proper and reproducable techniques. Many approaches rely on old or not perfectly suitable datasets because it results in huge effort to improve or create new datasets.

One has to note that this seminar paper only contains publicly available music recommendation systems. Outside of the research environment recommender systems are a very competitive field. Large companies such as the music provider service *Spotify*[5] have a desperate need of good music recommendation to satisfy their user's needs and to stand out from their competitors. A large share of their employees probably does similar work but is following a confidentiality agreement due to their contract.

4   Conclusions

Recommender systems is a very large and broadly spread research topic. With recent development in the field of artificial intelligence a lot of new approaches and ideas were published.
In this work we tried to summarize current developments as well as concerns evolving around methods and evaluation, especially for usage in a real-world environment.

There is a lot of work aiming to include new topics and relevant information into the task of recommendation. Often unfortunately a lot of limitations occur and methods are not really applicable for real-world usage.

---

[5] https://open.spotify.com/intl-de

To conclude we can say that there is a tendency to use novel methods such as deep learning to improve traditional approaches. Additionally there is a lot of research evolving around the incorporation of information from other disciplines such as for example the soziological or psychological effects.
We are probably not there yet but overall most of the recent publication aim to go step into the right direction.

References

1. Wang, Dongjing and Zhang, Xin and Yu, Dongjin and Xu, Guandong and Deng, Shuiguang: Came: Content-and context-aware music embedding for recommendation. In: IEEE Transactions on Neural Networks and Learning Systems, pp. 1375–1388 (2020)
2. La Gatta, Valerio and Moscato, Vincenzo and Pennone, Mirko and Postiglione, Marco and Sperlí, Giancarlo: Music recommendation via hypergraph embedding. In: IEEE transactions on neural networks and learning systems (2022)
3. Niyazov, Aldiyar and Mikhailova, Elena and Egorova, Olga: Content-based music recommendation system. In: 2021 29th Conference of Open Innovations Association (FRUCT), pp. 274–279 (2021)
4. Melchiorre, Alessandro B and Zangerle, Eva and Schedl, Markus: Personality bias of music recommendation algorithms. In: Proceedings of the 14th ACM Conference on Recommender Systems, pp. 533–538 (2020)
5. Priyanka, V Tejaswini and Reddy, Y Reshma and Vajja, Dharani and Ramesh, G and Gomathy, S: A Novel Emotion based Music Recommendation System using CNN. In: 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 592–596 (2023)
6. Singh, Jagendra and Sajid, Mohammad and Yadav, Chandra Shekhar and Singh, Shashank Sheshar and Saini, Manthan: A novel deep neural-based music recommendation method considering user and song data. In: 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), pp. 1–7 (2022)
7. Moscato, Vincenzo and Picariello, Antonio and Sperli, Giancarlo: An emotional recommender system for music. In: IEEE Intelligent Systems, vol. 36, pp. 57–68 (2020)
8. Hansen, Casper and Hansen, Christian and Maystre, Lucas and Mehrotra, Rishabh and Brost, Brian and Tomasi, Federico and Lalmas, Mounia: Contextual and sequential user embeddings for large-scale music recommendation. In: Proceedings of the 14th ACM Conference on Recommender Systems, pp. 53–62 (2020)
9. Cremonesi, Paolo and Jannach, Dietmar: Progress in recommender systems research: Crisis? What crisis?. In: AI Magazine, vol. 42, number 3, pp. 43–54 (2021)
10. Jannach, Dietmar and Zanker, Markus and Felfernig, Alexander and Friedrich, Gerhard: Recommender systems: an introduction. In: Cambridge University Press (2010)
11. Schafer, J Ben and Frankowski, Dan and Herlocker, Jon and Sen, Shilad: Collaborative filtering recommender systems. In: The adaptive web: methods and strategies of web personalization, pp. 291–324 (2007)
12. Burke, Robin: Hybrid recommender systems: Survey and experiments. In: User modeling and user-adapted interaction, vol. 12, pp. 331–370. Springer (2002)
13. Ferraro, Andres and Serra, Xavier and Bauer, Christine: Break the loop: Gender imbalance in music recommenders. In: Proceedings of the 2021 conference on human information interaction and retrieval, pp. 249–254 (2021)