MUSIC RECOMMENDATION USING AUDIO FEATURES

Boris Velichkov and Ivan Koychev

Departmet Software Technologies
Faculty of Mathematics and Informatics
Sofia University „St. Kliment Ohridski”, Bulgaria
boris.velichkov@gmail.com, koychev@fmi.uni-sofia.bg

***Abstract**: The paper presents a research that aims to investigate whether sound features can be used for recommending music. First it presents a study of existing tools for sound processing in order to see what features of the sound can be extracted with these tools. Second it presents experiments that use machine learning algorithms to identify the key features of the sound for the purpose of recommending music. Finally, manually classified data from 19 users were used for experiments. The achieved maximum average accuracy was measured to be 68.16%. This is an 18.17% increase in accuracy over the baseline. The conclusion is that it makes sense to analyze sound for the purpose of recommending music.*

***Keywords**: Music Information Retrieval, Recommender System, Audio Features.*

1. Introduction

With the advancement of technology and the growth of the information online, the access to any information also increases. Any multimedia, in particular music, is easily accessible to every user. Many systems offer access to different and varied musical performances. The question arises as to how the user can find the music he or she likes. Currently many systems successfully recommend music using standard approaches for recommendation. These approaches are based on metadata and user activity. In recent years, the processing and analysis of sound has been a popular research topic in the field of artificial intelligence. All this makes sound analysis a viable option for music recommendation. The aim of this paper is to verify whether analysis of sound and extraction of its features is useful for recommending music.

The remainder of this paper is structured as follows: Section 2 is a discussion of the related work. Section 3 explains the algorithm and the features used for music recommendation. Section 4 discusses the dataset, the choice of playlist, and audio feature extraction tools. This section also describes the classification process and presents the results of applying the algorithm to the dataset. Finally, Section 5 draws a conclusion.

2. Related Works

Spotify[1] is an online music service, which claims to uses sound features for music recommendations. They use deep convolutional neural networks to do this. As a result, they significantly outperform the more traditional approach using bag-of-words representations of audio signals [10].

Huang et al. (2004) compare three different methods: LBG vector quantization method, KNN and FAR classification algorithms. The results from their experiments provide evidences that the LBG vector quantization performs better than others and the feature of audio signature mean and variance are better predictors than audio signature raw data [4].

The project MusicSurfer uses a high-level music similarity metric to build a content-based music recommendation system. Main features are the rhythm and the timbre [2].

Bogdanov et al. (2011) use SVM for a content-based music recommendation. They automatically generate a semantic representation of the user's musical preferences directly from audio [1].

A collaborative music recommender system (CMRS) using audio features is presented in the paper of Li et al. (2006). The paper proposed a probabilistic model which adopts various clustering algorithms [5].

Another approach for music recommendation is using social tags. Social tagging is becoming increasingly popular in music information retrieval (MIR). It allows users to tag music items like songs, albums, or artists. Social tags are valuable to MIR, because they comprise a multifaceted source of information about genre, style, mood, users' opinion, or instrumentation [3, 9].

In our work, we present a classification approach that uses mostly low-level audio features. Related works were useful in choosing the algorithm used and the selection of the audio features.

3. Method for Content-based Music Recommendation

3.1. Feature extraction

In general for a Recommendation Systems, the used features are the most important factor for a correct result. The tool that extracts the audio features must be carefully selected in order to provide accurate data. A detailed research was made [11]. The paper of Moffat et al [8] was really helpful for the purposes of the research. jAudio [6, 7] was chosen because of its ease of use and the range of audio features it provides.

---

[1] https://www.spotify.com/

3.2. Machine Learning algorithms

To test our hypothesis whether music can be recommended on the basis of sound features we decided to use the k-nearest neighbors algorithm. The reasons for this choice are:

- This algorithm looks for closest neighbors, which means that it will found the nearest songs. This supposedly finds songs that most resemble those that the user likes [13].
- It is usually used for recommendation based on content (Content-based Filtering) [12].
- It is proposed as a good option. Although it is claimed that vector quantization yields better results [4].
- It is straightforward to configure and there are a lot of implementation available [13].
- K-nearest neighbors gave better results from Naive Bayes classifier on the initial configuration of the dataset.

During the experiments we tested different configurations of k-nearest neighbors. It was established that only changes in the parameter k have a significant impact on the resulting accuracy.

Based on the related literature and our research, we concluded that the different algorithms would yield similar results. Therefore, we concentrated our research on feature selection.

3.3. Feature Selection

The feature selection is made up of two main phases. The first phase is selecting a set of key features based on the related literature. Then, we applied an approach for feature selection based on the Apriori algorithm (choosing only those features that improve results) in order to reduce the number of features. The second phase starts with the best configuration of features obtained during the first phase. Then the same Apriori based approach is applied using all available features that jAudio provides. We used this process because it would not have been practicable to examine all possible configurations of features (244 combinations).

The ultimate goal of the experiments was to maximize the average accuracy for all users.

4. Experiments

4.1. Dataset

The data is one of the most important things in any task in the field of Machine Learning. If the data is not sufficient, not complete, or not consistent, contains wrong information, or is not handled in an appropriate form, it could lead to incorrect results.

To be able to check whether the sound characteristics can be used for recommending music, a user has to provide several things:

- A sufficient number of songs that can be analyzed and recommended.
- Users who will listen to and classify these songs.
- An audio features extraction tool.

Musical performances, which are used in the current work, are selected based on the ranking of billboard 100 best songs of all time[2]. What makes the selection of these songs good is that there is a high probability of each user to find the songs he/she likes. Also, this will facilitate the volunteers included in the experiment, because the music is available online for free via a Spotify account.

In order to check whether recommending songs based on musical features works, the volunteers need to listen to all 100 songs and note whether they like it or not. In practice they will classify the dataset. Volunteers are collected through a request for voluntary participation and are provided survey to fill out. The survey is a list of songs for which each volunteer must indicate whether he/she likes it or not on a per-song basis.

As a final result there are 19 datasets. Each dataset contains features derived by jAudio for the selected 100 songs and each song is classified by a volunteer. The classes are two - "yes" if the volunteer likes the song and "no" if he/she doesn't like it. In other words, the musical preferences of 19 users are recorded.

4.2. Results

Two best configurations of features were identified based on the conducted experiments.

The first configuration has the highest average accuracy for all users - 68.16%. This configuration also yields the maximum accuracy for an individual user - 91% and the minimum accuracy for an individual user - 50% (the green line in Figure 1).

In the second configuration there is a 2% lower average accuracy for all users - 66.16% accuracy. The maximum accuracy for an individual user is again 91%, but the minimum accuracy is 7% higher - 57% (the blue lines in Figure 1).

[2] http://www.billboard.com/charts/greatest-hot-100-singles

There is a 5.26% difference of accuracy between the best configuration of features (the green line) and the least performing one (the red line). It is an 18.17% increase in accuracy over the baseline.

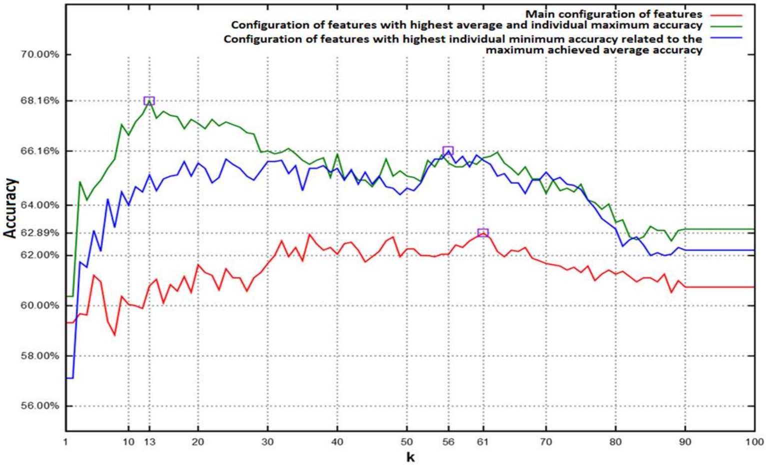

Fig 1. Chart showing the basic configuration of features, along with the two best configurations of features for all possible values of the parameter k.

The accuracy of the two best configurations of features is shown in Figure 2. It can be seen that the second configuration improves the results for the users that have been classified with the lowest accuracy.

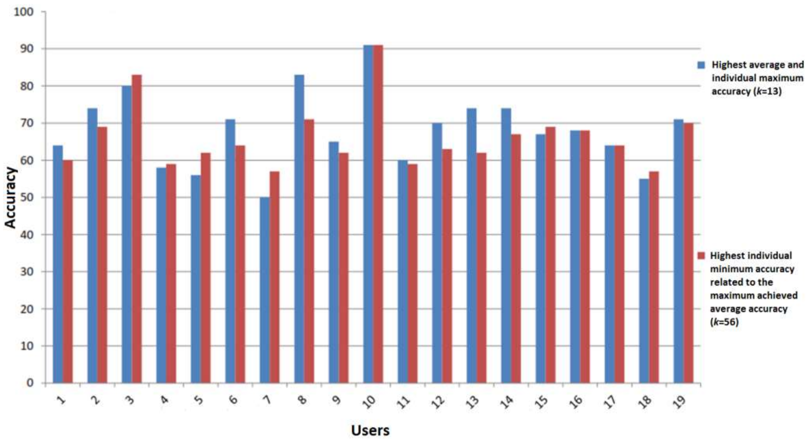

Fig 2. The accuracy of the two best configurations: the highest average and maximum accuracy depicted in blue and the highest individual maximum accuracy related to the maximum average accuracy depicted in red.

The best configuration, with 68.16% accuracy, has the features:

- Spectral Flux Overall Average
- Spectral Variability Overall Standard Deviation
- Compactness Overall Average
- Strength Of Strongest Beat Overall Standard Deviation

In the second best configuration with accuracy of 66.16%, the features are the same plus one more - Strongest Beat Overall Standard Deviation.

The relationship between all key metrics for the algorithm using the first configuration of features (68.16% accuracy) is shown in the Figure 3.

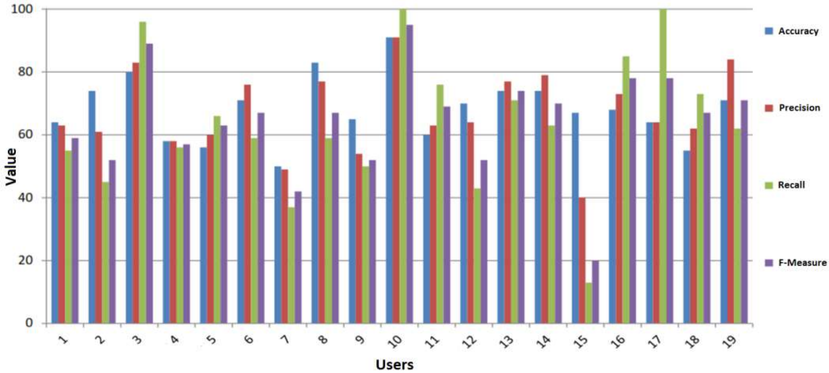

Fig 3. The results for each individual user in the configuration of the features with the highest average accuracy to all users (k=13)

The chart shows that there are two users with 100% recall. This means that the algorithm has returned all the songs they like.

Based on these results we can assume that if we try to optimize the algorithm for an individual user, we could get much better results.

5. Conclusion

During the experiments, the manually classified data of the 19 volunteers was used. The achieved maximum average accuracy was measured to be 68.16%. There is a 5.26% difference of accuracy between the best configuration of features and the least performing one. The best configuration scores an 18.17% increase in accuracy over the baseline. The average accuracy and the achieved maximum accuracy for an individual user prove that music features based on sound can be used for music recommendation. The most significant features are related to the spectrum, the compactness and the strongest beat. The conclusion that can be drawn is that it makes sense to analyze sound for the purpose of recommending music.

Acknowledgements

The work is supported by the NSF of Bulgaria under Grant No.: DN 02/11/2016 - ITDGate.

Bibliography

1. Bogdanov, Dmitry, et al. "Semantic audio content-based music recommendation and visualization based on user preference examples." Information Processing & Management 49.1 (2013): 13-33.
2. Cano, Pedro, Markus Koppenberger, and Nicolas Wack. "Content-based music audio recommendation." Proceedings of the 13th annual ACM international conference on Multimedia. ACM, 2005.
3. Eck, Douglas, et al. "Automatic generation of social tags for music recommendation." Advances in neural information processing systems. 2008.
4. Huang, Yao-Chang, and Shyh-Kang Jenor. "An audio recommendation system based on audio signature description scheme in mpeg-7 audio." Multimedia and Expo, 2004. ICME'04. 2004 IEEE International Conference on. Vol. 1. IEEE, 2004.
5. Li, Qing, Sung Hyon Myaeng, and Byeong Man Kim. "A probabilistic music recommender considering user opinions and audio features." Information processing & management 43.2 (2007): 473-487.
6. McKay, Cory, Ichiro Fujinaga, and Philippe Depalle. "jAudio: A feature extraction library." Proceedings of the International Conference on Music Information Retrieval. 2005.
7. McKay, Cory. "jAudio: Towards a standardized extensible audio music feature extraction system." Course Paper. McGill University, Canada (2005).
8. Moffat, David, David Ronan, and Joshua D. Reiss. "An evaluation of audio feature extraction toolboxes." Proceedings of the 18th International Conference on Digital Audio Effects (DAFx-15), Trondheim, Norway. 2015.
9. Nanopoulos, Alexandros, et al. "Musicbox: Personalized music recommendation based on cubic analysis of social tags." IEEE Transactions on Audio, Speech, and Language Processing 18.2 (2010): 407-412.
10. Van den Oord, Aaron, Sander Dieleman, and Benjamin Schrauwen. "Deep content-based music recommendation." Advances in neural information processing systems. 2013.
11. Velichkov, Boris, et al. “Classification, Recognition, Detection and Recommendation of Musical Performances”, Master Thesis
12. Wen, Zeng. "Recommendation System Based on Collaborative Filtering." CS229 Lecture Notes (2008).
13. Wu, Xindong, et al. "Top 10 algorithms in data mining." Knowledge and information systems 14.1 (2008): 1-37.