Music Genre Classification with FMA-LTS2 dataset

Ashutosh Mahala, Xiaosui Zhang

**Abstract**

Music genre classification problem is a challenging problem in machine learning which requires a large number of audio data as well as trustworthy labeling from music experts. In this paper, we use a novel dataset FMA-LTS2[1] (2017) which based on FMA [2] (Free Music Archive). We examine the FMA-LTS2 dataset carefully and point out the major advantages of using FMA-LTS2 compared to the other music classification dataset in the past. In the end, we run the FMA-LTS2 dataset through several standard machine learning algorithms and get the baseline of how these models performs.

1. **Introduction**

Internet and digital technologies have greatly changed the way Music Industry operates in the past decade. Due to the convenience of digital upload and download, individual music producers are able to publish their products much easier than their ancestors. Music industry becomes more decentralized and we’ve seen more tracks produced yearly [3]. Music genre classification is a process usually conducted by music experts in past. But with the rapid growth new tracks, the restriction of human reviews appears — human-reviewing are more time-consuming and sometimes hard to reach consensus.

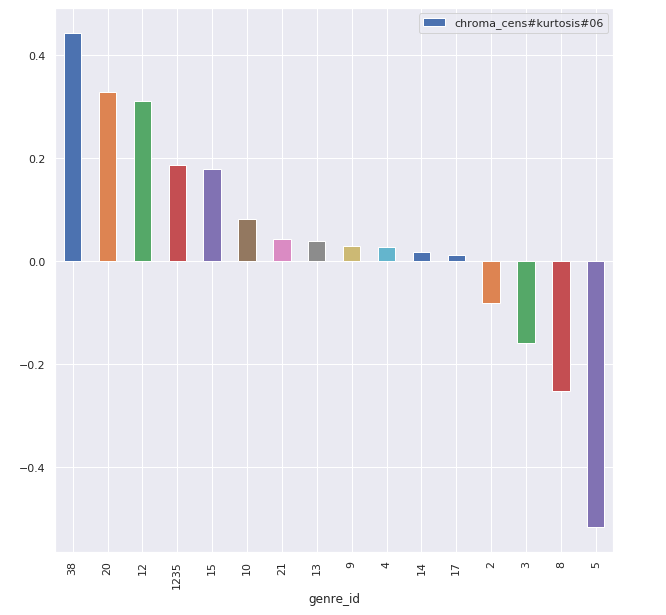
Many researches have been conducted aiming to solve this problem by using machine learnings algorithms to classify the music genre. Though the accuracy of these studies varies from 60% to 90% [8][9][10], the music data of these studies are either to biased like focusing on one language of songs or only have a very limit number of samples. However, in reality, the number of songs uploaded each day can be tremendous and very diverse in terms of language, genre and so on. In that case, A bigger dataset with less biased content will be crucial for the music genre classifiers to succeed in the reality. We will prove that FMA-LTS2 is such a dataset.

In second 2, we will look at each of acoustic features and explain the

1. **Data Analysis**

Free Music Archive(FMA) is a library of high quality songs, legal audio download directed by WFMU. Every music present on the FMA is precleared for certain types of uses otherwise restricted by copyright [2]. FMA dataset is the dataset which has been created from the songs from FMA. The main dataset size is 879GB which contains 106,574 full length tracks; it also provides a dataset with processed audio which is divided into 12 windows and then extracted the audio features out of those windows [1]. The feature total of 14 features have been extracted out which have been further divided in statistical methods and windows, thereby making it 518 features. This dataset contains 163 genres out of which we are only using the 16 genres.

Chroma is a feature which has not be well testified in the past. Chroma based features, which are corelated to harmony of the music. The phenomenon that human perception of pitch is periodic in the sense that two pitches are perceived as similar in “color” if they differ by an octave. Based on this observation, a pitch can be separated into two components, which are referred to as tone height and chroma [13].

Chroma\_cens is further degree of abstraction by considering short-time statistics over energy distributions within the chroma bands, one obtains CENS (Chroma Energy Normalized Statistics) features, which constitute a family of scalable and robust audio features. These features have turned out to be very useful in audio matching and retrieval applications [14, 15]. Chroma\_cens feature here shows the shows the general trend of difference in the tails octaves of particular types of genre. This could be a good feature in classification as the behaviors of different genre are seen very differently.

1. **Method**

We decided to apply Random Forest, RNN and SVM to the FMA-LTS2 dataset using the standard library from Sci-learn [100] and TensorFlow [101]. For each of the algorithm, we pick 3 hyperparameters for tuning and use a grid search to find the best combination based on the overall accuracy.

* 1. Random Forest

The three tuning hyperparameter of Random Forests are

* 1. SVM

SVM are known to

**Conclusion**

The efficiency and accuracy of music genre classifier relies heavily on the quality of input dataset. In this paper, we have shown that the FMA-LTS2 is high quality dataset which contains new acoustic features which hasn’t been properly examined in the past. We’ve also shown that the FMA-LTS2 dataset is less biased and covers more edge samples than the other datasets in the past. In the end, we applied three different machine learning models to the FMA-LTS2 dataset and achieves a highest 69% accuracy using SVM with linear kernel which provides a baseline for the future studies,

**REFERENCES**

[1] [Michaël Defferrard](https://arxiv.org/search/cs?searchtype=author&query=Defferrard%2C+M), [Kirell Benzi](https://arxiv.org/search/cs?searchtype=author&query=Benzi%2C+K), [Pierre Vandergheynst](https://arxiv.org/search/cs?searchtype=author&query=Vandergheynst%2C+P), [Xavier Bresson](https://arxiv.org/search/cs?searchtype=author&query=Bresson%2C+X), "FMA: A Dataset For Music Analysis", 2017

[2] Free Music Archive, http://freemusicarchive.org/

[3] Florida R, Jackson S, "Sonic city: the evolving economic geography of the music industry", 2008

[8] Min Xu; et al. (2004). “HMM-based audio keyword generation”

[9] Sahidullah, Md.; Saha, Goutam (May 2012). "Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition"

[10] George Tzanetakis, Perry Cook (2002) Musical Genre Classification of Audio Signals

[100] Scikit-learn, https://scikit-learn.org/stable/

[101] Tensorflow, <https://www.tensorflow.org/>