COMPUTING A SONG'S DECADE

Chris Latina

Georgia Tech Center for Music Technology
840 McMillan St. Atlanta, GA
chrisrlatina@gmail.com

Liang Tang

Georgia Tech Center for Music Technology
840 McMillan St. Atlanta, GA
liangmt@gatech.edu

ABSTRACT

Are there new insights through computational methods to the problem of getting song's year? This project, motivated by combining traditional music information retrieval (hereafter referred to as MIR) approach and metadata of song to compute a song's decade. Our task is to determine whether a heirarchical approach to classification using genre can help predict other metadata about a set of audio clips, such as the year of release. While genre classification is a common MIR problem, we set out to investigate whether one can use audio features in conjunction with metadata cross referenced with Discogs to estimate a song's decade.

Index Terms— Year computation, MIR, GTZAN

1. INTRODUCTION

MIR is an active field of machine learning research. People using online music services are very likely to search for music by genre, style or year, so researchers have attempted to understand how to automatically classify music by genre and style. For instance, overarching genres like rock or disco likely exhibit enough distinction for computers to effectively distinguish between them [1].

Many datasets that exist for genre classification only contain the minimum amount of information, namely the audio snippet and a label. We used a modified version of the GTZAN dataset augmented by metadata and preview clips form 7digital.com. After creating the list of files with the song name and artist, a set of scripts queries the discogs database and constructs a more complete dataset with Artist, Title, Album, Genre, Styles, and year of master release.

Discogs is an user-sourced database and marketplace for album releases on vinyl, cassette, CD, and digital formats. Unlike last.fm, the data is extremely complete and accurate because it is heavily moderated. Notably, all releases have a "master format", often chronologically the first release with metadata such as genre, styles (subgenres), and year.

2. RELATED WORK

2.1. Style in Music

Style is a vague term that may be understood in different ways depending on the context in which it is employed. Unlike the majority of other artistic disciplines, music is usually considered as not being able to generate creative work directly from the concrete reality [5]. As a result, style is slightly different in music than in literature, painting or sculpture. According to Dannenberg, it is almost impossible to find any obvious objective meaning, or referent associated with a short melody without words [6]. Essentially, every aspect of melody that communicates with listener is an aspect of style. One would say style is everything in music, or everything in music is style.

In regards to classification with metadata, "Computation Analysis Of Musical Influence" by Nick Collins mentions "...sourcing data from allmusic.com, and utilising python APIs for last.fm, EchoNest, and MusicBrainz." In his approach, Collins also cross referenced the year of the primary release with Discogs.com for accuracy. This paper touches upon the idea of a more semantic understanding of the output of classifiers.

In that effort, several papers have explored the efficacy of learning algorithms to predict genres. In his paper, George Tzanetakis effectively classified genres on live radio broadcasts using a Gaussian classifier [2]. Mandel used SVM on artist and album-level features to make similar classification as well [3]. Also, another study explored mixtures of Gaussians and K-Nearest-Neighbors (hereafter referred to as KNN) for music classification [4]. Each of these studies used similar features - Mel-Frequency Cepstral Coefficients (hereafter referred to as MFCC) and chroma features of audio to make the classification.

3. ALGORITHM OVERVIEW & DESCRIPTION

3.1. Audio Feature Extraction

3.1.1. Timbral Texture Features

MFCC are perceptually motivated features that are based on the STFT.

$$\mathrm{MFCC}(n) = \sum_{k'=1}^{\mathcal{K}'} \log\left(|X'(k',n)|\right) \cdot \cos\left(j \cdot \left(k' - \frac{1}{2}\right) \frac{\pi}{\mathcal{K}'}\right)$$

The Spectral Centroid (hereafter referred to as SC) is defined as the center of gravity of the magnitude spectrum of the STFT.

$$\mathrm{SC}(n) = \frac{\sum_{k=0}^{\mathcal{K}/2-1} k \cdot |X(k,n)|^2}{\sum_{k=0}^{N/2-1} |X(k,n)|^2}$$

The centroid is a measure of spectral shape and higher centroid values correspond to ”brighter” textures with more high frequencies.

3.1.2. Instantaneous Feature

The Spectral Flux (hereafter referred to as SF) measures the amount of change of the spectral shape. It is defined as the average difference between consecutive STFT frames: (should be an equation here).

$$\mathrm{SF}(n,\beta) \quad = \quad \frac{\sqrt[\beta]{\sum_{k=0}^{\mathcal{K}/2-1} \left(|X(k,n)| - |X(k,n-1)|\right)^{\beta}}}{\mathcal{K}/2}$$

3.2. Dataset Creation

In this research, GTZAN was chosen as the basic database. It is a database of music created by George Tzanetakis specifically for machine learning analysis of genre classification problems. Though, this dataset was argued to have non-negligible issues such as ”album effect”, and repeated audio samples which would negatively influence the accuracy of the classification [7], it still can be seen as the default choice in the research area of music genre classification. Figure 1 represents the degree of activity of GTZAN.

Our dataset started as 1000 songs from 10 genres but many of the audio clips lacked information, were duplicate songs, or entire albums. In an attempt to maintain accuracy, reduce the album effect, and distribute the years represented, we reduced the dataset to 7 genres with 60 songs in each genre. The majority of these are from GTZAN with some instances added from 7digital.com. We used the 7digital API endpoint interface to search within genres and scrape the 30 second preview clip. Although some of the previews from 7digital represent remastered versions, we cross referenced with the Discogs master release for consistency. Ideally, the dataset would have an even distribution of songs spanning the range of all years without repeated artists.

The final dataset consists of 420 audio clips in total with genres Blues, Country, Disco, Hip Hop, Metal, Pop, Rock. Any additional track samples added from 7digital were cut to 30 seconds in length using ffmpeg. Metadata was then scraped from Discogs.com API using python.

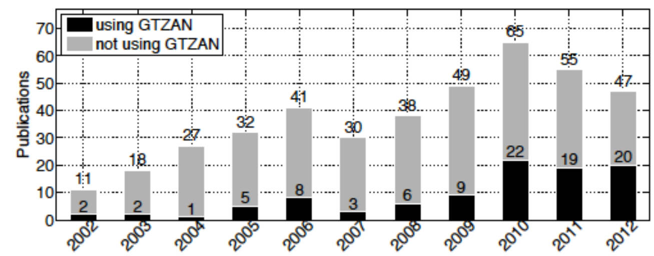

Fig. 1. Annual numbers of published works in MGR with experimental components, divided into ones that use and do not use GTZAN [7]

3.3. Query Metadata

Discog.....

3.4. Classification

For classification, two approaches which are based off of Support Vector Machine (hereafter referred to as SVM) and KNN were used. The idea of applying two different approaches to achieve year computation is not only trying to maximize the accuracy of year computation, but comparing the suitability of this two basic statistical pattern recognition (hereafter referred to as SPR) classifiers. Essentially, SPR is applied to estimate the probability density function (hereafter referred to as PDF) for the feature vectors of each class.

4. EVALUATION

4.1. Methodology

4.1.1. N-fold Validation

N-fold was applied in the research in order to perform multiple iteration of classification for evaluation. Given to the fact that the dataset contains 8 genres with 60 songs each (420 songs in total), N was set to 7. This way, for each fold, the training set contains 350 songs and the test set contains 70 songs, with 10 from each genre.

4.1.2. Segmentation of Data

The data was segmented into training and testing pools using two techniques. The results were quite different and tie into some fundamental concepts of dataset creation. One approach was an interleaved distribution. The second was a randomized distribution.

For the interleaved distribution, the feature set was read in order adding one song from each genre at a time. Feature set was read from a tab-delimted file containing the song's metadata including artist name, album, title, genre, year and filename ordered by genre, artist, and then album alphabetically . Because there were instances of multiple songs by an artist from the same album, the distributed segmentation approach separates songs from one album occurring in both the training and test set, thus removing the risk of the "artist effect", "album effect". In his paper, Seyerlehner argues that it is relatively easy to identify songs by one and the same artist using audio similarity algorithm [8]. This effect is kown as artist effect. In some cases, even album-specific production dffects are reflected in the spectral representation of songs, which is respectively called "album effect". Obviously, songs by the same artist will tend to belong to the same genre, and the ability to recognize the genre by specific production effect is not what this research intend to measure. However, this also the worst case scenario, because the training and test set will always be very different from each other.

For the randomized distribution, a seed is generated for the entire feature set and applied to all songs and metadata, completely randomizing the order. This creates a potentially even distribution of genres within each fold for the training and test sets. It greatly increases the chance of the album effect, however, because songs from the same album that were previously only in the training set now can appear in the test set as well.

4.2. Metrics

5. CONCLUSION

5.1.

5.2.

6. FUTURE WORK

6.1. Feature Extraction

6.2. Classification

6.3. Data set enhancement

7. REFERENCES

[1] Omar Diab, Anthony Manero, and Reid Watson, "Musical genre tag classification with curated and crowd-sourced datasets," .

[2] George Tzanetakis and Perry Cook, "Musical genre classification of audio signals," *Speech and Audio Processing, IEEE transactions on*, vol. 10, no. 5, pp. 293–302, 2002.

[3] Michael I Mandel and Daniel PW Ellis, "Song-level features and support vector machines for music classification," in *ISMIR 2005: 6th International Conference on Music Information Retrieval: Proceedings: Variation 2: Queen Mary, University of London & Goldsmiths College, University of London, 11-15 September, 2005*. Queen Mary, University of London, 2005, pp. 594–599.

[4] Tao Li, Mitsunori Ogihara, and Qi Li, "A comparative study on content-based music genre classification," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*. ACM, 2003, pp. 282–289.

[5] Hugo Leichtentritt, "Aesthetic ideas as the basis of musical styles," *The Journal of Aesthetics and Art Criticism*, vol. 4, no. 2, pp. 65–73, 1945.

[6] Roger B Dannenberg, "Style in music," in *The structure of style*, pp. 45–57. Springer, 2010.

[7] Bob L Sturm, "The gtzan dataset: Its contents, its faults, their effects on evaluation, and its future use," *arXiv preprint arXiv:1306.1461*, 2013.

[8] Klaus Seyerlehner, Gerhard Widmer, and Tim Pohle, "Fusing block-level features for music similarity estimation," in *Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10)*, 2010, pp. 225–232.