Genre Classification of Music by Tonal Harmony

Carlos Pérez-Sancho CPEREZ@DLSI.UA.ES
David Rizo DRIZO@DLSI.UA.ES
Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante, Spain

Stefan Kersten SKERSTEN@IUA.UPF.EDU
Rafael Ramirez RRAMIREZ@IUA.UPF.EDU
Music Technology Group, Universitat Pompeu-Fabra, Barcelona, Spain

Keywords: genre classification, language models, chord transcription, chord progressions

Abstract

We present a genre classification framework for audio music based on a symbolic classification system. Audio signals are transformed to a symbolic representation of harmony using a chord transcription algorithm, by computing Harmonic Pitch Class Profiles. Then, language models built from a groundtruth of chord progressions for each genre are used to perform classification. We show that chord progressions are a suitable feature to represent musical genre, as they capture the harmonic rules relevant in each musical period or style.

1. Genre classification

Organization of large music repositories is a tedious and time-intensive task for which music genre is an important meta-data. Automatic genre and style classification have become popular topics in Music Information Retrieval (MIR) research because musical genres are categorical labels created by humans to characterize pieces of music and this nature provides the genre meta-data with a high semantic and cultural information to the music items in the collection.

Traditionally, the research domain of genre classification has been divided into the audio and symbolic music analysis and retrieval domains. Nevertheless, some authors have paid attention recently on making use of the best of both worlds. The work by Lidy et al. (Lidy et al., 2007) deals with audio to MIDI transcription in order to extract features from both signals and then combine the decisions of the different classifiers. On the other hand, Cataltepe and coworkers' approach (Cataltepe et al., 2007) is just the opposite: to synthesize audio from MIDI and then analyze both signals to integrate the classifications.

Our proposal is to use tonal harmonic information to distinguish between musical genres. The underlying hypothesis is that each musical genre makes use of different rules that allow or forbid specific chord progressions. As it can be found in (Piston, 1987), some rules that were almost forbidden in a period have been accepted afterwards. Also, it is well known that pop-rock tunes mainly follow the classical tonic-subdominant-dominant chord sequence, whereas jazz harmony books propose different series of chord progressions as a standard.

The goal of this work is to classify digital audio music using a language modelling system trained from a groundtruth of chord progressions, bridging the gap between audio and symbolic by means of a chord transcription algorithm.

2. Chord transcription

The Pitch Class Profile (PCP) measure has been used in automatic chord recognition or key extraction since its introduction by Fujishima (Fujishima, 1999). The perception of musical pitch has two main attributes: height and chroma. Pitch height moves vertically in octaves telling which octave a note belongs to, while chroma tells its position in relation to others within an octave. A chromagram or a pitch class profile is a 12-dimensional vector representation of a chroma, which represents the relative intensity in each of twelve semitones in a chromatic scale. Since a chord is composed of a set of tones, and its label is only determined by the position of those tones in a chroma, regardless of their heights, chromagram seems to be an ideal feature to represent a musical chord.

Table 1. Classification accuracy percentages using different n-gram lengths.

Data set	2-grams	3-grams	4-grams
3 classes	54.4	61.1	61.1
popular vs. jazz	70.9	83.4	83.4
academic vs. jazz	75.0	75.0	75.0
academic vs. popular	56.7	66.7	66.7

In this paper we have obtained the Harmonic Pitch Class Profile (HCPC) by applying the algorithm in (Gómez & Herrera, 2004), which deviates from Fujishima's PCP measure by distributing spectral peak contributions to several adjacent HCPC bins and taking a peaks harmonics into account. The feature vectors are then processed to obtain a symbolic representation of chords in the form of triads. Thus, each song is represented as a string of chord progressions, with an alphabet of 24 symbols (major and minor triads).

3. Language models

We have used a language modelling approach to classify chord progressions (Camastra & Vinciarelli, 2008). The idea is to capture the different use of chord progressions in each style by using n-grams of chords taken from a training set. Thus, a language model is built for each genre in the training set, and these models are evaluated against a new set of songs. These songs are then labelled with the genre of the model with the highest probability of having generated it.

4. Experiments

Two different data sets were used in the experiments. As a training set we used a groundtruth of chord progressions, obtained from a set of 761 symbolic music files belonging to three genres: academic (comprising baroque, classical and romanticism), jazz and popular. For testing the system we used a set of 12 audio files from the same genres, from which we obtained the chord progressions by applying the chord transcription algorithm described in section 2.

Table 1 shows the classification accuracy of the system when using different n-gram sizes and configurations of the data sets (all-against-all and binary classifiers). As expected, better results were obtained when classifying jazz music against the other two styles. This is consistent with the fact that the differences in harmony between these styles are usually bigger than they are between academic and popular music. Also, results tend to improve when using larger n-grams, as they are more capable to capture the typical structures present in each style.

5. Conclusions and future work

We have shown the feasibility of classifying digital audio music by genre using a symbolic classification system. For this, a chord transcription algorithm was used to obtain the chord progressions present in each song. The results show a good performance when classifying between harmonically distant genres, while it is more difficult to distinguish when they make use of more similar chord progressions. Nevertheless, we are convinced that these results can be improved by using a richer chord vocabulary, to better represent the slight differences in harmony that are lost when using only chord triads. Also, a bigger set of audio files will be constructed to deeply test this approach.

Acknowledgments

This work is supported by the Spanish PROSEMUS project (TIN2006-14932) an partially supported by the research programme Consolider Ingenio 2010 (MIPRCV, CSD2007-00018).

References

Camastra, F., & Vinciarelli, A. (2008). *Machine learning for audio, image and video analysis*. Springer. 1st. edition.

Cataltepe, Z., Yaslan, Y., & Sonmez, A. (2007). Music genre classification using midi and audio features. *EURASIP Journal on Advances in Signal Processing, 2007*.

Fujishima, T. (1999). Realtime chord recognition of musical sound: A system using common lisp music. *Proceedings of the International Computer Music Conference*. Beijing.

Gómez, E., & Herrera, P. (2004). Automatic extraction of tonal metadata from polyphonic audio recordings. *Proceedings of 25th International AES Conference*. London, UK.

Lidy, T., Rauber, A., Pertusa, A., & Iñesta, J. (2007). Improving genre classification by combination of audio and symbolic descriptors using a transcription system. *Proceedings of the ISMIR* (pp. 61–66). Vienna, Austria.

Piston, W. (1987). *Harmony*. Norton, W. W. & Company, Inc. 5th. edition.