Music Genre Classification

Single-labelled Music Genre Classification Using Content-based Features

Ritesh Ajoodha

Degree of Doctor of Philosophy in Computer Science
The University of the Witwatersrand
Supervised by Dr. B. Rosman; Mr. R. Klein; and Prof. E. Momoniat

Contents

Introduction

- Genre is the **most common** classification scheme used to distinguish music
- There exists a consensus of broad genre definitions **across populations** worldwide
- Similarity-based measures include mood, artist, and style.
- Genre offers a **culturally authorised prominence** on the construction of traditional classes

Research Motivation I

1. Music recommendation
2. Music information retrieval
3. Musicological significance

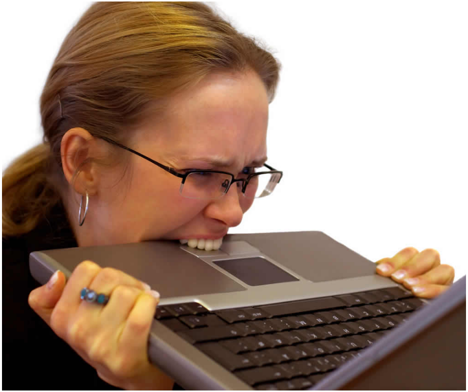
"Its kind of fun to do the impossible." - *Walt Disney*

Problem Statement I

- Traditional musical aspects given by four characteristics: melody, harmony, rhythm and sound.
- Textbook definitions are qualitative and come across as subjective, context dependent and therefore are difficult to automate.
- Composers do not abide by "genre definitions"
- Humans often **cognitively** regard art and other manifestations of genre collectively (e.g. food, cloths, language, artwork, music), which could bias the study.
- Since people always disagree with what a particular genre is, correct classification becomes inescapably bounded
- Genre holds many sub-genres
- There is an awareness of genre classification performance bounds imposed by humans.

Problem Statement II

- Humans are biased and subjective in their classifications
- Genre definitions evolve



Problem Statement III

Literature Review

There are very few capable genre classification systems

Benetos and Kotropoulos [2008]	75.0%
Bergstra et al. [2006]	82.5%
Holzapfel and Stylianou [2008]	74.0%
Li et al. [2003]	79.7%
Lidy et al. [2007]	76.8%
Panagakis et al. [2008]	78.2%
Sturm [2013]	83.0%
Tzanetakis and Cook [2002]	61.0%

Table: Classification of 10-GTZAN genre

Contributions

Briefly stated, we provide the following contributions

- A thorough review of music genre classification literature
- Features and classification algorithms never used for genre classification
- A list of features that best distinguishes different genres
- Detailed comparison of representations to build an optimal classifier

Feature Analysis

- Acoustic content comprises of instrument sounds, speech sound, and environmental sounds.
- Human listeners try to identify these characteristics to classify a piece of music
- Four main categories:
  1. The Magnitude Spectrum
  2. Tempo Detection
  3. Pitch Detection
  4. Chordal Progressions

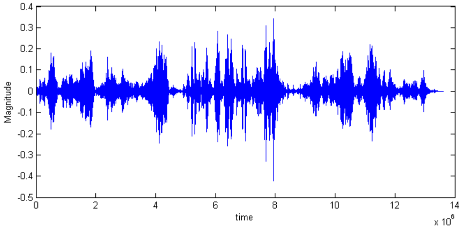
Feature Representations

We used the followiong representations:

1. **Central Tendency:** Mean and standard deviation.
2. **The Feature Histogram:** The feature histogram arranges the features local window intensities into bin ranges
3. **MFCC Aggregation:** MFCC representation is a wellknown feature representation which takes the first n MFC coefficients of the feature samples as it would a 16khz signal.
4. **Area Moments:** Image moments is a central concept in computer vision and has its root in image processing.

Magnitude-based Features I

- The magnitude spectrum, obtained from the fast Fourier transform of a signal, houses a family of spectral features for genre classification.
- We can now Identify signal change, noisiness, loudness and many other spectral features.
- Exploring peak-based features allows us to analyse the signal more thoroughly.



Magnitude-based Features II

Some of the magnitude features include:

1. Spectral Slope
2. Compactness
3. Spectral Decrease
4. Loudness
5. Onset Detection
6. Peak Detection
7. Spectral Flux
8. Spectral Variability
9. Mel-Frequency Cepstral Coefficients
10. Spectral Flatness

Magnitude-based Features III

Result Preview

149 features from the magnitude spectrum were extracted

Feature	Optimal Representation	Dimensionality
Slope	Mean	1
Compactness	Mean	2
Decrease	Mean	1
Loudness	Mean	26
Onset Detection	Mean	1
Octave Based Signal Intensity	Mean	17
Peak-based features	Mean	4
Spectral Flux	MFCC	4
Spectral Variability	MFCC	4
MFCC	MFCC	52
Flatness	Mean	20
Shape Statistics	Mean/MFCC	11
Spectral Rolloff	Mean	2
Peak Flux	20-bin FH	20
Crest Factor	Mean	10
Strongest Freq of FFT Max	MFCC	4

Table: Magnitude-based feature list

Tempo Features I

- Most music display regular rhythmic formation that creates an impression of tempo.

Tempo Features II

Some of the magnitude features include:

1. Energy
2. Beat Histogram

Tempo Features III

Result Preview

362 tempo related features were extracted

Feature	Optimal Representation	Dimensionality
Energy	Mean	2
Fraction of low energy	Mean	2
Beat Histogram	Mean	342
Strongest Beat	Mean	2
Strength of the Strongest Beat	Mean	2
Beat Sum	MFCC	4
Relative Difference Function	MFCC	4
Temporal Shape Statistics	Mean	4

Table: Tempo feature list

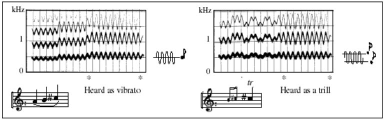
Pitch Features I

- Pitch is a perceived characteristic contained in the frequency of music content.
- Most music of the same genre exhibit melodies that are just combined notes from a scale set.
- However, often environmental sounds overtone pitch, disguising available pitch related elements, which make it difficult to extract pitch computationally.
- Therefore, some sort of pitch extraction mechanisms need to be adopted to retrieve these pitch elements though the environmental sounds.

Pitch Features II

Some of the magnitude features include:

1. Amplitude Modulation
2. Zero Crossing Rate



Pitch Features III

Result Preview

75 pitch related features were extracted

Feature	Optimal Representation	Dimensionality
Autocorrelation Coefficients	Mean	49
Amplitude Modulation	Mean	8
Zero Crossing	MFCC	4
Envelope Statistics	Mean	4
LFS	Mean	10

Table: Pitch feature list

Chordal Features I

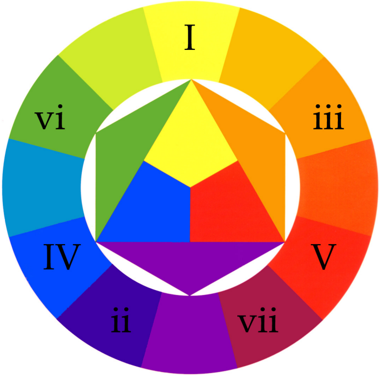
- Introducing spectral feature extraction to genre detection problems created opportunities to exploit single characteristics of music.
- Chord structure and progressions has been a defining trait of music for many years but had gone unnoticed in recent music genre detection schemes.

I	ii	iii	**IV**	**V**	vi	vii
Major	Minor	Minor	**Major**	**Major**	Minor	Dim.
A	B	C#	**D**	**E**	F#	G#
B	C#	D#	**E**	**F#**	G#	A#
C	D	E	**F**	**G**	A	B
D	E	F#	**G**	**A**	B	C#
E	F#	G#	**A**	**B**	C#	D#
F	G	A	**Bb**	**C**	D	E
G	A	B	**C**	**D**	E	F#

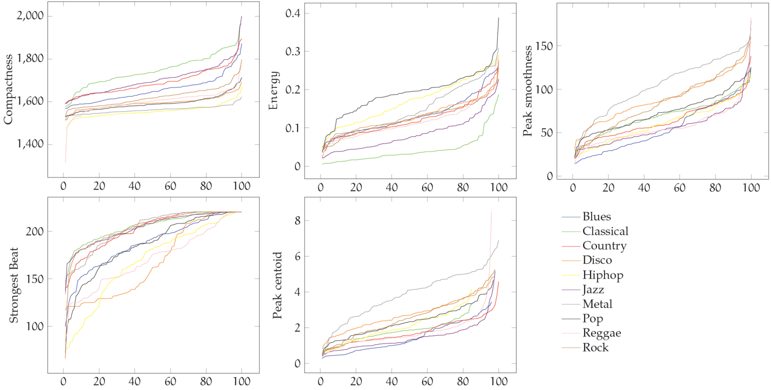
Chordal Features II

Chordal Progressions

1. Chroma



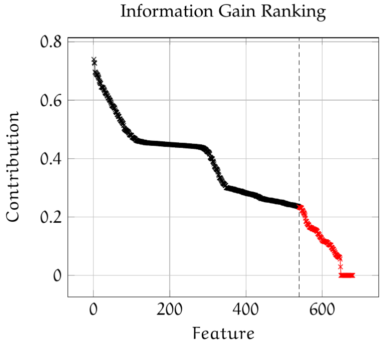
Graphical Overview of Features



Feature Selection I

Feature Selection

We chose the first 539 features



Feature Selection II

Information Gain Ranking

We could have chosen up to 100 features and achieved between 70-75% classification accuracy, but doing this would bias the learning model to this particular dataset.

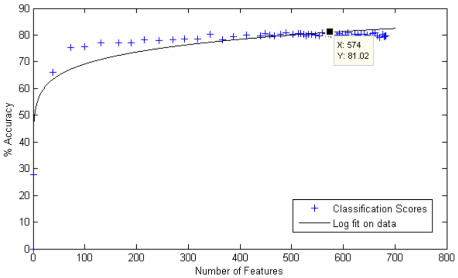


Figure: Feature vs Classification Accuracy

Feature Selection III

Features Maintained (459)
Spectral Flux (MFCC 4)
Spectral Variability (MFCC 4)
Compactness (Mean + SD 2)
MFCCs (MFCC 52)
Peak Centroid (Mean + SD 2)
Peak Smoothness (SD 1)
Complex Domain Onset Detection (Mean 1)
Loudness (+ Sharpness and Spread) (Mean 26)
OBSI (+ Radio) (Mean 17)
Spectral Decrease (Mean 1)
Spectral Flattness (Mean 20)
Spectral Slope (Mean 1)
Shape Statistic spread (Mean 1)
Spectral Centroid (MFCC 4)
Spectral Rolloff (SD 1)
Spectral Crest (Mean 19)
Spectral Variation (Mean 1)
Autocorrelation Coefficients (Mean 49)
Amplitude Modulation (Mean 8)
Zero Crossing + SF (MFCC 8)
Envelope Statistic Spread (1)
LPC and LSF (Mean 12)
RMS (Mean + SD 2)
Fraction of Low Energy (Mean 1)
Beat Histogram (SD) (171)
Strength of Strongest Beat (Mean 1)
Temporal Statistic Spread (Mean 1)
Chroma (MFCC 48)

Features Eliminated (223)
Peak Flux (20-bin FH 20)
Peak Smoothness (Mean 1)
Shape Statistic centroid, skewness and Kurtosis
Strongest Frequency of Centroid (MFCC 4)
Spectral Rolloff (Mean 1)
Strongest Frequency of FFT (MFCC 4)
Envelope Centroid, Skewness and Kurtosis (Mean 4)
Beat Histogram (Mean 171)
Strongest Beat (Mean + SD 2)
Strength of Strongest Beat (SD 1)
Fraction of Low Energy (SD 1)
Beat Sum (MFCC 4)
Relative Difference Function (MFCC 4)
Temporal Statistic Centroid, Skewness & Kurtosis

Automatic Music Genre Classification

Classifying 10 GTZAN genres

- Although the multilayer perception takes a significant time to build and evaluate, it notably outperforms the naïve Bayes and the SVM.
- The K-NN and RF take the least time to build and evaluate and produce sufficient results.
- The LLRM provides the best classification score.

Classifier	Accuracy	Time to build model	Time to evaluate model
Naïve Bayes	46.40%	0.11 sec	2.13 sec
Support vector machines	32.50%	6.04 sec	38.12 sec
Multilayer perceptron	63.70%	635.37 sec	6 hours 20.12 sec
Linear logistic regression models	81.00%	20.25 sec	10 mins 31 secs
K-nearest neighbours	72.80%	0.02 sec	13.12 sec
Random forests	66.60%	0.22 sec	3.76 sec

Table: Classification of the thinned feature vector

Linear Logictic Regression Classification

LLRM 10-fold CV on 10-GTZAN Genres

G1 = Blues, G2 = Classical, G3 = Country, G4 = Disco, G5 = Hiphop, G6 = Jazz, G7 = Metal, G8 = Pop, G9 = Reggae, and G10 = Rock.

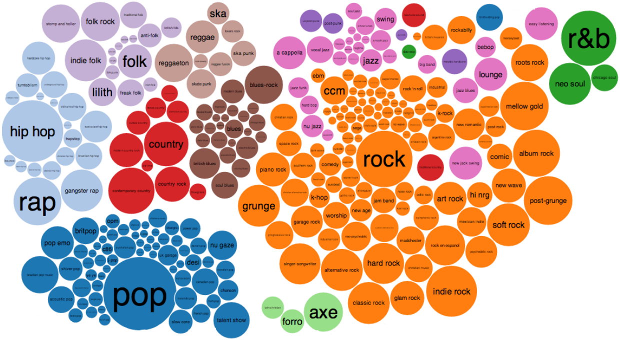
	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8	G_9	G_{10}
G_1	84	0	3	3	0	5	1	0	2	2
G_2	0	96	1	0	0	2	0	0	0	1
G_3	3	0	77	2	0	4	0	1	3	10
G_4	1	1	5	76	2	0	0	4	5	6
G_5	1	0	0	1	85	0	4	3	6	0
G_6	3	4	5	1	0	82	1	2	1	1
G_7	2	0	0	1	1	0	90	0	0	6
G_8	0	0	4	4	1	0	0	84	1	6
G_9	2	0	3	6	6	1	1	4	70	7
G_{10}	5	0	7	9	2	0	5	5	1	66

Conclusion and Future Work I

- Humans who observe cultural features not content based features
- Large scale musical structures are present in most music genre types (RNNs)
- Compile quality datasets with masterful labelling
- Construct datasets based on different characteristics that a genre learning model should exhibit and detect
- The musicality of a listener can be be used to satisfy a particular customer's genre preference
- The pallet of genre labels used
- Current genre labelling places albums and artists into genre catalogues.
- A promising approach is performing multi-label automatic classification, which offers a solution to the fuzziness between genre definitions.

Conclusion and Future Work II

- Understanding the strength of connections between genre tags could aid the construction of a weighted graph that maps genre to genre relationships.
- The type of mis-classifications that a learning model can achieve.



References and Additional Reading

1. Ritesh Ajoodha, *Automatic Music Genre Classification.* The University of the Witwatersrand.
2. Adam Berenzweig, Beth Logan, Daniel PW Ellis, and Brian Whitman. *A large-scale evaluation of acoustic and subjective music-similarity measures.* Computer Music Journal, 28(2):63-76, 2004.
3. Dannenberg, Roger B., Belinda Thom, and David Watson. "*A machine learning approach to musical style recognition.*" (1997).
4. Deshpande, Hrishikesh, Rohit Singh, and Unjung Nam. *"Classification of music signals in the visual domain."* Proceedings of the COST-G6 Conference on Digital Audio Effects. sn, 2001.
5. Li, Tao, Mitsunori Ogihara, and Qi Li. *"A comparative study on content-based music genre classification."* Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval. ACM, 2003.
6. Silla, Carlos N., Celso AA Kaestner, and Alessandro L. Koerich. *"Automatic music genre classification using ensemble of classifiers."* Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on. IEEE, 2007.
7. Liu, Dan, Lie Lu, and HongJiang Zhang. *"Automatic mood detection from acoustic music data."* ISMIR. 2003.
8. McKay, Cory, and Ichiro Fujinaga. *"Automatic Genre Classification Using Large High-Level Musical Feature Sets."* ISMIR. Vol. 2004. 2004.
9. Chen, Ling, Phillip Wright, and Wolfgang Nejdl. *"Improving music genre classification using collaborative tagging data."* proceedings of the second ACM international conference on web search and data mining. ACM, 2009.
10. Jiang, Dan-Ning, et al. *"Music type classification by spectral contrast feature."* Multimedia and Expo, 2002. ICME'02. Proceedings. 2002 IEEE International Conference on. Vol. 1. IEEE, 2002.
11. Tzanetakis, George, and Perry Cook. *"Musical genre classification of audio signals."* Speech and Audio Processing, IEEE transactions on 10.5 (2002): 293-302.

Acknowledgments

Thanks

Benjamin Rosman

Richard Klein

Mike Mchunu

NRF National Research Foundation

UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG

GOLDEN KEY INTERNATIONAL HONOUR SOCIETY · 1977 ·