MUSICAL GENRE CLASSIFICATION Combining Features Extracted From Audio, Symbolic, and Cultural Sources

Burcak Agridag burcak.agridag@rwth-aachen.de

June 29, 2016



Contents



Introduction

- Different Data Sources
- Accuracy Improvement
- History
- Methodology
 - Audio Features
 - Audio Features
 - Symbolic Features

4 Experiments

- Experiment Results
- 5 Conclusion
 - Demo & Questions

Introduction

Musical Genres

- A musical genre is characterized by the common characteristics shared by its members as a result of a complex interaction between the public, marketing, historical, and cultural factors.
- These characteristics typically are related to the instrumentation, rhythmic structure, and harmonic content of the music.

Genre Classification







Source Types

- Audio (e.g. MP3)
- Symbolic (e.g. MIDI)
- Cultural (e.g. web data, user tags, surveys)
- 7 possible combinations

Audio

- Audio processing is the fundamental way in which the music is consumed.
 - Spectral, time domain information extracted directly from audio(not intuitively musical).
- Symbolic
 - High-level musical features can be extracted easily compared to mere audio data.
 - Polyphonic audio to symbolic transcription algorithms continue to improve.
 - Instruments, Chord Progressions, Pitches, Rythmic Patterns, etc..

Cultural

- Labeling musical pieces based on past experience and knowledge primarily shaped by the culture.
 - Composer, Style, Genre(!), etc.

Classification Accuracy rate(%)	3-Resources	1-Resource
5 genre taxonomies	96.8%	85.5%
10 genre taxonomies	78.8%	65.1%



- Focus on either audio, symbolic, or cultural sources on information.
- Significant research on audio-cultural data, Whitman and Smaragdis [1].
- Less work on audio-symbolic data, Lidy et al. [2].
- Significantly increased performance when cultural features are used [3].

Methodology

Methodology Overview

• jAudio: Audio feature extractor.



• jSymbolic: Symbolic feature extractor.

$$\mathbf{MIDI} \rightarrow \left(\begin{array}{c} \mathbf{111} \\ \mathbf{features} \end{array}\right)$$

• jWebMiner: Cultural feature extractor.

OF HIT COUNTS FOR DIFFERENT SEARCH STRINGS (KEYWORDS)

- 26 core features (e.g. Zero Crossing Rate, MFCCs, Spectral Flux, etc.)
- 5 meta-features to combine with core features (Mean, Derivative of the mean, Standard Deviation, etc.)
- 3 aggregators: Mean, Standard Deviation, and MFCCs (output on per-song basis instead of per window basis).

• e.g. Spectral Flux, spectral correlation btw. adjacent windows as an indication of the degree of change of the spectrum:

$$F_t = \sum_{n=1}^{N} \left(N_t[n] - N_{t-1}[n]
ight)^2$$

• e.g. RMS, calculate the amplitude of a window:

$$RMS = \sqrt{\frac{\sum_n^N x_n^2}{N}}$$

• e.g. MFCCs.



- Low-level features, spectral or time-domain information:
 - Spectral flux, RMS, zero-crossing rate, ...
- High-level features, musical abstractions:
 - Instruments, Melodic Contour, ...
- Cultural Features, sociocultural information:
 - Playlist co-occurrence, purchase correlations,

Instrumentation:

- What types of instruments are present?
- Which are given particular importance?, ..
- Rhythm:
 - Time intervals between the attacks of different notes
 - Durations of each note,..
- Pitch statistics:
 - Occurrence rates of different notes, ...

- Given Primary, Secondary search strings, and required filter words search the Web for all possible combinations:
 - e.g. "Beethoven" + "Classical" + "musician"
 - e.g. "Beethoven" + "Metal" + "musician"
 - Count the # of hits for each occurrence of those 3 together and hold statistics.

Experiments

• The SAC (Symbolic, Audio and Cultural) dataset consists of 250 MIDI files and 250 matching MP3s, accompanying metadata (e.g., title, artist, etc.) in iTunes XML file format.

Blues: Modern Blues and Traditional Blues Classical: Baroque and Romantic Jazz: Bop and Swing Rap: Hardcore Rap and Pop Rap Rock: Alternative Rock and Metal

Figure: 10 genres in SAC dataset.

Feature Type	5-Genre Code	10-Genre Code
Symbolic	S-5	S-10
Audio	A-5	A-10
Cultural	C-5	C-10
Symbolic + Audio	SA-5	SA-10
Audio + Cultural	AC-5	AC-10
Symbolic + Cultural	SC-5	SC-10
Symbolic + Audio + Cultural	SAC-5	SAC-10

- For each of the 10-genre experiments, a normalized weighted classification accuracy rate was calculate to provide insight on error types:
 - Misclassification within a genre pair (e.g., Alternative Rock instead of Metal):

0.5 x ERROR

• Misclassification outside a genre pair (e.g., Swing instead of Metal):

1.5 x ERROR

Figure: Classification accuracy (%)

	S	Α	С	SA	AC	SC	SAC
5-	86.4	82.8	87.2	92.4	95.2	94	96.8
UW							
10-	66.4	67.6	61.2	75.6	78.8	75.2	78.8
UW							
10-	66.4	67.4	66.6	78.6	84.6	81.2	84.2
W							

	1 Feature Type	2 Feature Types	3 Feature Types
5-UW	85.5	93.9	96.8
10-UW	65.1	76.5	78.8
10-W	66.8	81.5	84.2

Figure: Classification accuracy (%)





- For automatic genre classification task, it is beneficial to combine features extracted from audio,symbolic and cultural data sources.
- Combining feature types decreased the seriousness of those mis-classifications, particularly when cultural features are included.

Demo & Questions

Demo & Questions

?

[1] McEnnis, D., C. McKay, and I. Fujinaga. 2006. jAudio: Additions and improvements. Proceedings of the International Conference on Music Information Retrieval. 385–6.

[2] McKay, C., and I. Fujinaga. 2006. jSymbolic: A feature extractor for MIDI files. Proceedings of the International Computer Music Conference. 302–5.

[3] McKay, C., and I. Fujinaga. 2007. jWebMiner: A web-based feature extractor. Proceedings of the International Conference on Music Information Retrieval. 113–4.

 http://music.ece.drexel.edu/files/genreJawn.png.
 http://www.informatik.uniaugsburg.de/de/lehrstuehle/hcm/projects/internal/mmgc/img/mmgclogo.j
 http://albumware.ru.
 ccrma.stanford.edu.

JAudio Feature Extractor File Edit Recording Analysis Playback Help

RECORDINGS:			FE	AT	TURES:			
Name	Path			30/8	Feature		Dimension	15
p0051disc1side1_16bit_44100hz	/home/soul/public_html/handel/m		17	-	Magnitude Spectrum		variable	٦.
p0051disc1side2_16bit_44100hz	/home/soul/public_html/handel/m				Power Spectrum		variable	
lp0051disc2side3_16bit_44100hz	/home/soul/public_html/handel/m				FFT Bin Frequency Labels		variable	
lp0051disc2side4_16bit_44100hz	/home/soul/public_html/handel/m			r	Spectral Centroid		1	
lp0336disc1side1_16bit_44100hz	/home/soul/public_html/handel/m				Derivative of Spectral Centroid		1	
lp0336disc1side1_16bit_44100hz	/home/soul/public_html/handel/m				Running Mean of Spectral Centroid		1	ъ
lp0336disc1side2_16bit_44100hz	/home/soul/public_html/handel/m				Standard Deviation of Spectral Centro	id	1	
lp0337disc1side1_16bit_44100hz	/home/soul/public_html/handel/m				Derivative of Running Mean of Spectra	al Centroid	1	
lp0337disc1side1_16bit_44100hz	/home/soul/public_html/handel/m				Derivative of Standard Deviation of Sp	pectral Centroid	1	
lp0337disc1side1_16bit_44100hz	/home/soul/public_html/handel/m			r	Spectral Rolloff Point		1	
lp0337disc1side2_16bit_44100hz	/home/soul/public_html/handel/m				Derivative of Spectral Rolloff Point		1	
lp0337disc1side2_16bit_44100hz	/home/soul/public_html/handel/m				Running Mean of Spectral Rolloff Point	t	1	
lp0337disc1side2_16bit_44100hz	/home/soul/public_html/handel/m				Standard Deviation of Spectral Rolloff	Point	1	
Ip0402disc1side1_16bit_44100hz	/home/soul/public_html/handel/m				Derivative of Running Mean of Spectra	al Rolloff Point	1	
lp0402disc1side1_16bit_44100hz	/home/soul/public_html/handel/m				Derivative of Standard Deviation of Sp	pectral Rolloff Point	1	
lp0402disc1side1_16bit_44100hz	/home/soul/public_html/handel/m		H	~	Spectral Flux		1	
lp0402disc1side1_16bit_44100hz	/home/soul/public_html/handel/m				Derivative of Spectral Flux		1	
lp0402disc1side1_16bit_44100hz	/home/soul/public_html/handel/m				Running Mean of Spectral Flux		1	
lp0402disc1side2_16bit_44100hz	/home/soul/public_html/handel/m				Standard Deviation of Spectral Flux		1	
lp0402disc1side2_16bit_44100hz	/home/soul/public_html/handel/m				Derivative of Running Mean of Spectra	al Flux	1	
lp0402disc1side2_16bit_44100hz	/home/soul/public_html/handel/m		H		Derivative of Standard Deviation of Sp	pectral Flux	1	
lp0402disc1side2_16bit_44100hz	/home/soul/public_html/handel/m			r	Compactness		1	
lp0402disc1side2_16bit_44100hz	/home/soul/public_html/handel/m				Derivative of Compactness		1	
lp0402disc1side2_16bit_44100hz	/home/soul/public_html/handel/m				Running Mean of Compactness		1	
lp0418disc1side1_16bit_44100hz	/home/soul/public_html/handel/m	-			Standard Deviation of Compactness		1	
Add Recordings	Delete Recordings			s	ave Features For Each Window 🛛 🕒	Save For Overall F	Recordings	
Feature Values Save Path:	feature_values_1.xml		w	inc	low Size (samples): 5	12		
Feature Definitions Save Path:	feature_definitions_1.xml		w	inc	low Overlap (Traction): 0.	0		
						Extract Fea	atures	-

Figure: jAudio GUI

-			
Burcal		aric	20
Duica	$\sim -$	VELIC	I a E

🖆 jSymbolic Feature Extractor						_ 0 🛛	
RECORDINGS:			FEATURES	8			
Name	Path		Save	Feature		Dimensions	
day_fool.mid	C:\MIDI_Files\day_fool.mid	_	~	Melodic Thirds		1	
dc-choochoochaboogie.mid	C:\MIDI_Files\dc-choochoochaboogie			Melodic Tritones		1	
deadend.mid	C:\MIDI_Files\deadend.mid	1	×	Minimum Note Duration		1	
DeadMan.mid	C:\MIDI Files\DeadMan.mid	1	Image: A start and a start	Most Common Melodic Interva	1	1	
deb3.mid	C3MIDI Files\deb3.mid	1		Most Common Melodic Interva	I Prevalence	1	
deb5.mid	C:\MIDI_Files\deb5.mid	1	Image:	Most Common Pitch Class		1	
december.mid	C3MIDI_Files\december.mid	1		Most Common Pitch Class Pr	evalence	1	
dedi.mid	C3MIDI_Files\dedi.mid	ш.	~	Most Common Pitch		1	
Desafinado78.mid	C:\MIDI_Files\Desafinado78.mid		~	Most Common Pitch Prevalen	0.0	1	
despafm.mid	C:\MIDI_Files\despafm.mid	1	Image: A start and a start	Note Density		1	
despaw.mid	C3MIDI Files\despayy.mid	1	Image: A start and a start	Number of Common Melodic I	ntervals	1	
despdpc.mid	C3MIDI Files\despdpc.mid	1		Number of Common Pitches		1	
despeie.mid	C3MIDI Files\despeie.mid	1	~	Number of Moderate Pulses		1	
despmpl.mid	C:\MIDI_Files\despmpl.mid	1	Image: A start and a start	Number of Pitched Instrument	5	1	
dexterity.mid	C:\MIDI_Files\dexterity.mid	1	~	Number of Relatively Strong P	ulses	1	
dgaria22.mid	C3MIDI Files\dgaria22.mid	1	Image: A start and a start	Number of Strong Pulses		1	
DinahWashington-TroubleInMind.mid	C:\MIDI Files\DinahWashington-Trou	1		Number of Unpitched Instrum	ents	1	
dniazzme.mid	C3MIDI Files3dnjazzme.mid	1	Image: A start and a start	Orchestral Strings Fraction		1	
doc-whos been talkin.mid	C3MIDI Files\doc-whos been talkin	1		Overall Dynamic Range		1	
dock of the bay 2.mid	C3MIDI Files\dock of the bay 2.mid	1 1	~	Percussion Prevalence		1	
Doin Time.mid	C3MIDI Files\Doin Time.mid	1 1	~	Pitch Class Variety		1	
dongio04_madamina.mid	C:\MIDI_Files\dongio04_madamina	1	~	Pitch Variety		1	
donnalee.mid	C3MIDI Files\donnalee.mid	1		Polyrhythms		1	
dontmean.mid	C:\MIDI Files\dontmean.mid	1	Image: A start and a start	Primary Register		1	
Dose.mid	C3MIDI Files\Dose.mid	-	~	Quality		1	
Add Recordings	Validate Recordings		🕑 Do Not	Use Windows	Save Featu	res For Each Window	
	Store Sequence				Save For O	verall Recordings	
Delete Recordings			Window L	ength (seconds):			
			Window C	Window Overlap (fraction):			
Play Sequence	quence Stop Playback		Fea	Feature Values Save Path:		feature_values_1.xml	
	Feature Definitions Save Path:		ure Definitions Save Path:	feature_definitions_1.xml			
View File Info.					Ex	tract Features	

Figure: jSymbolic GUI

Appendix

🏂 jWebMiner							_ 🗆 🖂
Search Information	tion						
Search Words	Filter Words	Site Weightings	Options R	esults			
		Co-Occurre	nce Extraction	Cross Tabulat	tion Extraction		
		0 00 000000	ioo Entraotion	0 01000 10000	ion End dotton		
PRIMARY SEARCH	I STRINGS:			SECONDARY SEA	RCH STRINGS:		
Load	Save	Clear	Organize	Load	Save	Clear	Organize
FBigr John Patton HEREtoday. 10,000 Maniacs 2Pac 5D Cent Abole All Silmani Abdelhalim Hafez Abdolvahhab Sha Abelardo Barroso Abu Hial Abelardo Barroso Abu Hial Abelardo Barroso Abu Hial Abelardo Barroso Abu Hial Abrashi Affricando Adrica Bailey Aerosmith Adricando Afric Cuban All-Sta Agustín Carbonell Aid Ayubi Al Green Al Hibbler	est hidi "Sola"		×	Atternative Pop / F Blues - Contemp Blues - Country B Blues - Urban Blu Classical - 20th C Classical - 20th C Classical - 20th C Classical - Class Classical - Roma Country Dance Pop Electronica Mp Hop / Rp Jazz - Acid Jazz Jazz - Anarh Gard Jazz - Bebop Jazz - Col Jazz Jazz - Dixieland Jazz - Latin Jazz Jazz - Latin Jazz Jazz - Latin Jazz Jazz - Swing	Rock Iues Lues entury Classical entury Classical ical ssance & Med. nntic		A
			EXTRACT	FEATURES			

Figure: jWebMiner GUI



Figure: Classification



Figure: jMIR