Feature Extraction for Musical Genre Classification MUS15

Kilian Merkelbach

June 22, 2015

Intro -

Contents









What are genres?



Genres provide order

inde Musi	Musik Horen Events Charts k bel Last.fm k bel kastatalog (similar and a Tan anatata) (Similar)	Registrieren Einloggen
		Erhalte Musikempfehlungen ganz nach deinem Geschmack.
Alle	Bei Last.fm beliebte Musik	Melde dich an, sag uns welche Musik du magst und wir werden sofort neue Musik für dich forden
acoustic	Dopulär i Angesent i Depulär in Deutschland	werden solor neue Masik für üten inden.
ambient	Popular Angesagt Popular in Deutschland	Melde dich jetzt an
blues	Rihanna	Completion film in C
classical	154.971.550 Mal gespielt (4.180.591 Hörer)	scrobbler fur iOs
country	Ahnlich wie: Nicki Minaj, Beyoncé, Jessie J, Rita Ora, Ciara	Ein neuer iOS-Musikplayer, mit dem o
electronic	e pop	SC Musik wiederentdecken kannst, die d
emo		genort. Erfahre mehr
folk		
gothic	Coldplay 322.402.245 Mal gesplett (5.052.684 Hörer)	
hardcore	Ahnlich wie: Keane, OneRepublic, Snow Patrol, U2, Imagine Dragons	
hip hop	- rock	
indie	Constant of the	
industrial		
azz	AUE TOUR Mumford & Sons	
metal	79.374.192 Mal gespielt (1.559.511 Hörer)	
рор	Ahnlich wie: Of Monsters and Men, The Lumineers, The Wedding Band, Passenger, Jerry Douglas	
punk rock		

Intro -

Pipeline - Feature extraction



Intro -

Pipeline - Training and Classification



History -

Contents









History -

MP3 revolution

- 1998: First MP3 players available
- 1999: Napster, mp3.com go online
- Commercial and private music collections explode

History -

First steps

- 2001: Tzanetakis and Cook build foundation by publishing first paper
- from then on: methods evolve from previous ones, always trying to improve performance

Methods -

Contents







Methods



Methods -

State of the Art

- Tzanetakis, Cook: "Musical Genre Classification of Audio Signals"
- Correa, Costa, Saito: "Tracking the Beat: Classification of Music Genres and Synthesis of Rhythms"

Features Overview

- Timbral Texture Features (19 dim.)
- Rhythmic Content Features (6 dim.)
- Pitch Content Features (5 dim.)

Timbral Texture Features I

- based on short time Fourier transform (STFT)
- Analysis Window (23ms) vs. Texture Window (1s)



Timbral Texture Features II

- calculate four features from STFT output and use their mean and variance for classification
 - ► e.g. Spectral Flux:

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

 additional features based on Mel-frequency cepstral coefficients (MFCCs)

Rhythmic Content Features

- Discrete Wavelet Transform (DWT) to analyze signal
- autocorrelation function recognizes strong beats (example)
- map peaks into Beat Histogram

Beat Histogram



Pitch Content Features

 Pitch Histogram: like Beat Histogram (0.5-1.5s) but with shorter time frame (2-50ms)

Results

- 59% accuracy with 10 different genres
- 77% among classical genres
- 61% among jazz genres



AUDIO CLASSIFICATION HIERARCHY

Tracking the Beat

- Songs given as MIDI files
- Use directed graphs to describe song

Digraphs

- Weighted directed graphs with note lengths as vertices
- Weights defined by frequency of note sequence



Features

- Build digraph for each song: $18 \cdot 18 = 324$ dim.
- Additional features from digraph: 15 dim.
- Use PCA to obtain 52 dim. feature vector

Results

 85.72% accuracy with 4 different genres (blues, bossa-nova, reggae and rock) Conclusion -

Contents









Prediction by humans

- College students found correct genre (out of 10) with 70% accuracy after 3 seconds [3]
- Machines are on par
- Human experts still better if there are many genres

Future Research

- step away from speech recognition features (e.g. MFCCs)
- fuzzy classification (e.g. 90% Rock, 10% Blues)
- larger genre sets, higher accuracy

References



Wikimedia Commons.

Short time fourier transform, 2006.

- Debora C Correa, Luciano da F Costa, and Jose H Saito. Tracking the beat: Classification of music genres and synthesis of rhythms.
- G. Tzanetakis and P. Cook.

Musical genre classification of audio signals. Speech and Audio Processing, IEEE Transactions on, 10(5):293–302, Jul 2002.