A Multi-Layer Modeling Approach to Music Genre Classification

MC886 - Final Project

Renan V. Novas Gabriel S. Vicente

Universidade Estadual de Campinas

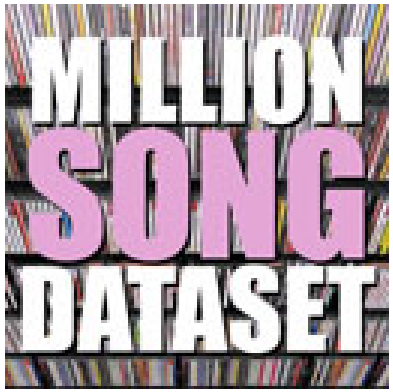
December 17, 2014

- Music Information Retrieval
  - Instrument and musical structure recognition
  - Recommendation and taste prediction engines
  - Automatic music transcription and algorithmic composition
  - Automatic categorization

last.fm

- **Ogihara, Mitsunori et al.** A Comparative Study on Content-based Music Genre Classification
- **Liang, Dawen et al.** Music Genre Classification with the Million Song Dataset

Million Song Dataset



- 1,000,000,000 tracks dataset!
  - The Echo Nest (Spotify) and LabROSA (Columbia University)

Purposes:

- To encourage research on algorithms that scale to commercial sizes
- To provide a reference dataset for evaluating research
- As a shortcut alternative to creating a large dataset with APIs
- To help new researchers get started in the MIR field

Million Song Dataset
Example Track

loudness	float	overall loudness in dB
mode	int	major or minor
mode confidence	float	confidence measure
release	string	album name
release 7digitalid	int	ID from 7digital.com or -1
sections confidence	array float	confidence measure
sections start	array float	largest grouping in a song, e.g. verse
segments confidence	array float	confidence measure
segments loudness max	array float	max dB value
segments loudness max time	array float	time of max dB value, i.e. end of attack
segments loudness max start	array float	dB value at onset
segments pitches	2D array float	chroma feature, one value per note
segments start	array float	musical events, ~ note onsets
segments timbre	2D array float	texture features (MFCC+PCA-like)
similar artists	array string	Echo Nest artist IDs (sim. algo. unpublished)
song hotttnesss	float	algorithmic estimation
song id	string	Echo Nest song ID
start of fade out	float	time in sec
tatums confidence	array float	confidence measure
tatums start	array float	smallest rythmic element
tempo	float	estimated tempo in BPM
time signature	int	estimate of number of beats per bar, e.g. 4

Recreating the dataset

- Issues
  - Unbalancedness
  - Tags subjectivity

Genre	Training	Test
dance and electronica	2,880	320
folk	2,880	320
jazz and blues	2,880	320
punk	2,880	320
soul and reggae	2,880	320
	14,400	1,600

- 30 audio-descriptive features inspired by The Echo Nest
- Referente features
  - Loudness
  - Tempo
  - Time signature
  - Key
  - Mode
  - Duration
- Timbral features
  - 12 timbral averages
  - 12 timbral variances

SVM classifiers

- One $\times$ One
- One $\times$ Rest

Confusion Matrix

Table: Confusion Matrix

	Predicted Genre					
	A	B	C	D	E	
A	224	21	24	17	34	320
B	12	224	39	15	30	320
C	27	42	223	9	19	320
D	24	15	7	240	34	320
E	39	36	16	11	218	320
	326	338	309	292	335	

Model Comparison