Dmitry Bogdanov, Alastair Porter | Music Technology Group, Universitat Pompeu Fabra
Julián Urbano | Multimedia Computing Group, Delft University of Technology
Hendrik Schreiber | tagtraum industries incorporated
Contact: dmitry.bogdanov@upf.edu

The MediaEval 2017 AcousticBrainz Genre Task

Content-based Music Genre Recognition from Multiple Sources

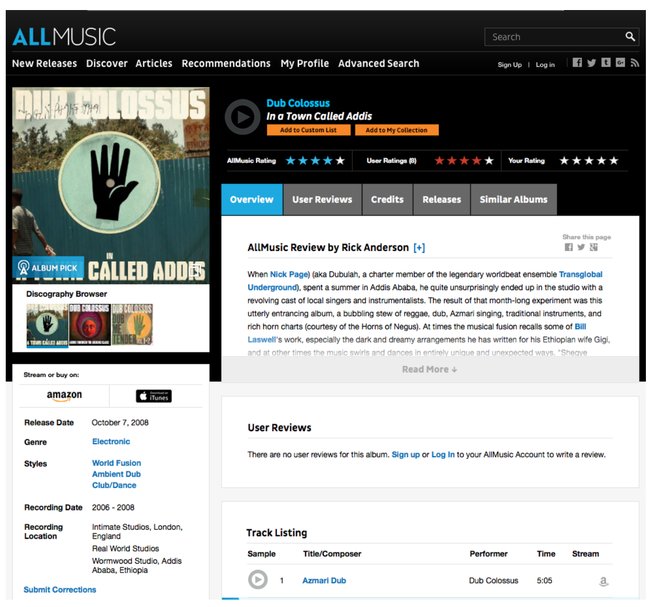
AcousticBrainz

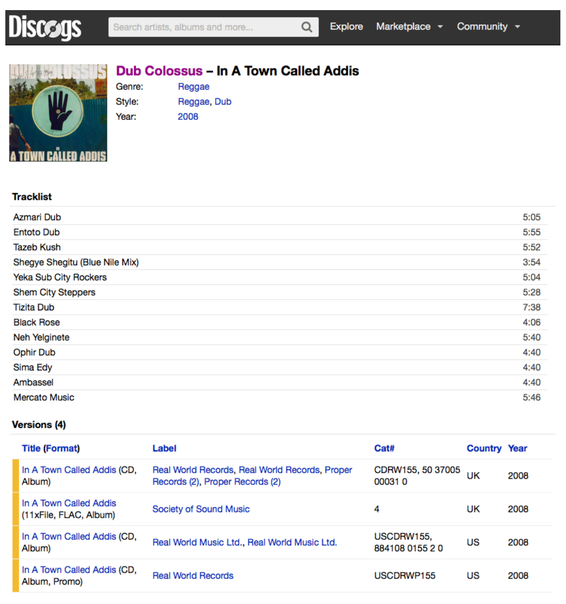
https://acousticbrainz.org

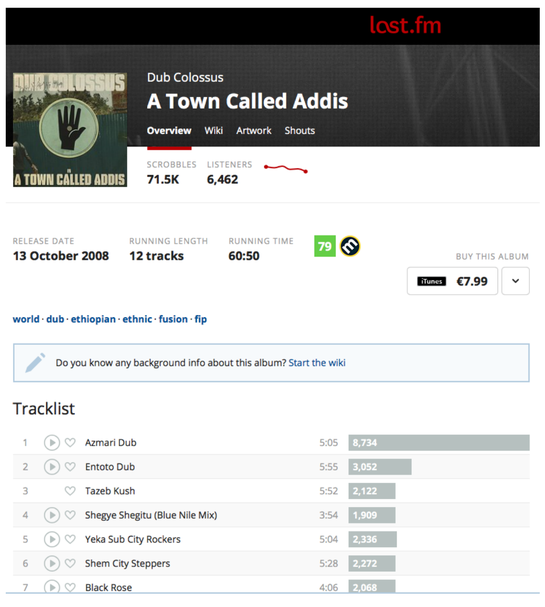
Community database of music features extracted from audio

- Open data computed by open algorithms (Essentia Music Extractor)
  http://essentia.upf.edu/documentation/streaming_extractor_music.html
- Built on submissions from the community
- Over 5,600,000 analyzed recordings (tracks)
- ~3000 music features (bags-of-frames)
- Statistical information about timbre, rhythm, tonality, loudness, etc.
- Rich music metadata from MusicBrainz

The problem of genres







Development datasets

Dataset	AllMusic	Discogs	Lastfm	Tagtraum
Type	Explicit	Explicit	Tags	Tags
Annotation level	Release	Release	Track	Track
Recordings	1,353,213	904,944	566,710	486,740
Release groups	163,654	118,475	115,161	69,025
Genres	21	15	30	31
Subgenres	745	300	297	265
Genres/track	1.33	1.37	1.14	1.13
Subgenres/track	3.15	1.69	1.28	1.72

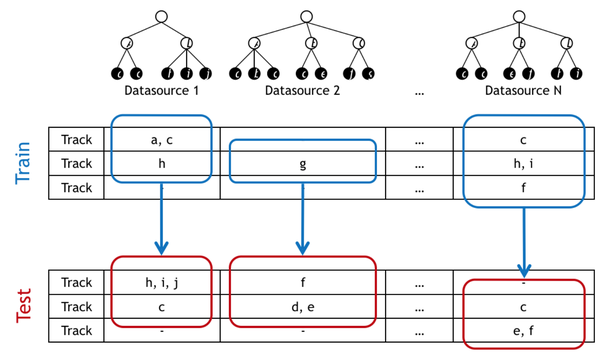
AcousticBrainz Genre Task

Goal: Predict genre and subgenre of unknown music recordings given precomputed music features

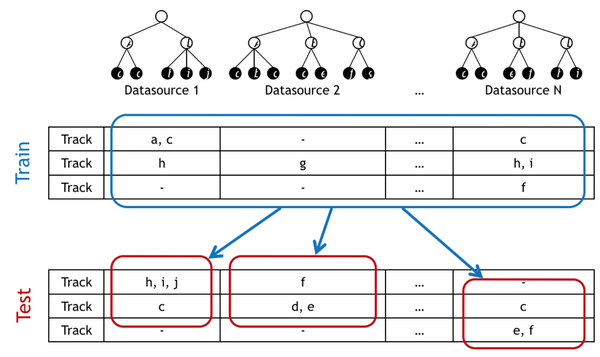
Task novelty:

- Four different genre annotation sources (and taxonomies)
- Hundreds of specific subgenres
- Multi-label genre classification problem
- A very large dataset (~2 million recordings in total)

Subtask 1:
Single-source Classification
Build a separate system for each ground-truth dataset



Subtask 2:
Multi-source Classification
Can we benefit from combining ground truths into one system?



Evaluation

Metrics: Precision, Recall and F-score

- Per recording, all labels (genres and subgenres)
- Per recording, only genres
- Per recording, only subgenres
- Per label, all recordings
- Per genre label, all recordings
- Per subgenre label, all recordings

Baselines

- Random baseline: following the distribution of labels
- Popularity baseline: always predicts the most popular genre

Reproducibility

- Open data
- Open-source code
- Music features extraction (Essentia)
- Genre metadata mining (MetaDB)
- Task evaluation and baselines

Submissions

- Participants from five teams
- Maximum 5 submissions for each task per team
  (5 submissions x 2 tasks x 4 datasets = 40 runs)
- 115 runs received in total

Results

Per-recording F-score (all labels)

ALLMUSIC				DISCOGS				LASTFM				TAGTRAUM			
SIGNIFICANTLY BETTER THAN BASELINES															
jku	T1	R2	0.442	jku	T1	R2	0.550	jku	T1	R2	0.556	jku	T1	R2	0.556
jku	T1	R1	0.431	jku	T1	R1	0.528	jku	T1	R1	0.546	jku	T1	R1	0.543
jku	T2	R4	0.431	jku	T2	R4	0.528	jku	T2	R4	0.546	jku	T2	R4	0.543
jku	T1	R3	0.426	jku	T1	R3	0.513	jku	T1	R3	0.546	jku	T1	R3	0.528
jku	T2	R5	0.426	jku	T2	R5	0.513	jku	T2	R5	0.546	jku	T2	R5	0.528
jku	T2	R2	0.326	dbis	T2	R5	0.426	jku	T2	R2	0.434	jku	T2	R2	0.432
				jku	T2	R2	0.400	jku	T2	R3	0.368	dbis	T2	R5	0.401
				dbis	T1	R1	0.398	dbis	T2	R4	0.366	dbis	T2	R4	0.386
				dbis	T2	R1	0.398	dbis	T2	R5	0.366	dbis	T1	R1	0.369
				dbis	T1	R2	0.390	dbis	T1	R3	0.340	dbis	T2	R1	0.369
				dbis	T2	R2	0.390	dbis	T1	R4	0.340	dbis	T1	R3	0.363
				dbis	T2	R4	0.382	dbis	T1	R5	0.340	dbis	T1	R4	0.363
				dbis	T1	R3	0.374	dbis	T2	R3	0.340	dbis	T1	R5	0.363
				dbis	T1	R4	0.374	jku	T2	R1	0.319	dbis	T2	R3	0.363
				dbis	T1	R5	0.374	dbis	T1	R1	0.318	dbis	T1	R2	0.349
				dbis	T2	R3	0.374	dbis	T2	R1	0.318	dbis	T2	R2	0.349
				jku	T2	R1	0.310	dbis	T1	R2	0.290	jku	T2	R1	0.345
								dbis	T2	R2	0.290	jku	T2	R3	0.311
SIMILAR TO POPULARITY BASELINE															
kart	T1	R3	0.257	kart	T1	R4	0.284	kart	T1	R3	0.188	kart	T1	R3	0.285
jku	T2	R3	0.254	kart	T1	R3	0.280	baseline	T1	R2	0.184	kart	T1	R4	0.251
jku	T2	R1	0.253	jku	T2	R3	0.278	kart	T1	R4	0.182	baseline	T1	R2	0.250
kart	T1	R4	0.250	baseline	T1	R2	0.269	kart	T1	R1	0.169	kart	T1	R1	0.180
baseline	T1	R2	0.249					kart	T1	R2	0.140				
dbis	T1	R3	0.249												
dbis	T1	R4	0.249												
dbis	T1	R5	0.249												
dbis	T2	R3	0.249												
kart	T1	R1	0.223												
SIMILAR TO RANDOM BASELINE															
dbis	T2	R5	0.183	samirit	T1	R1	0.166	samirit	T1	R1	0.107	samirit	T1	R1	0.165
samirit	T1	R1	0.181	kart	T1	R1	0.160	baseline	T1	R1	0.070	kart	T1	R2	0.133
dbis	T2	R4	0.176	kart	T1	R2	0.149	icsi	T1	R3	0.064	baseline	T1	R1	0.128
dbis	T1	R1	0.151	baseline	T1	R1	0.131	icsi	T1	R4	0.058	icsi	T1	R3	0.078
dbis	T2	R1	0.151	icsi	T1	R3	0.057	icsi	T1	R5	0.048	icsi	T1	R4	0.077
dbis	T1	R2	0.129	icsi	T1	R4	0.051	icsi	T1	R2	0.019	icsi	T1	R5	0.059
dbis	T2	R2	0.129	icsi	T1	R5	0.040	icsi	T1	R1	0.013	icsi	T1	R2	0.025
kart	T1	R2	0.120	icsi	T1	R2	0.026					icsi	T1	R1	0.022
baseline	T1	R1	0.110	icsi	T1	R1	0.021								

Per-label F-score (all labels)

ALLMUSIC				DISCOGS				LASTFM				TAGTRAUM			
SIGNIFICANTLY BETTER THAN BASELINES															
jku	T1	R1	0.196	jku	T1	R3	0.252	jku	T1	R1	0.351	jku	T1	R3	0.282
jku	T2	R4	0.196	jku	T2	R5	0.252	jku	T2	R4	0.351	jku	T2	R5	0.282
jku	T1	R3	0.196	jku	T1	R1	0.245	jku	T1	R3	0.349	jku	T1	R1	0.274
jku	T2	R5	0.196	jku	T2	R4	0.245	jku	T2	R5	0.349	jku	T2	R4	0.274
jku	T2	R2	0.191	jku	T1	R2	0.232	jku	T1	R2	0.346	jku	T1	R2	0.260
jku	T1	R2	0.179	jku	T2	R2	0.226	jku	T2	R2	0.316	jku	T2	R2	0.254
jku	T2	R3	0.170	jku	T2	R1	0.191	jku	T2	R3	0.292	jku	T2	R3	0.220
jku	T2	R1	0.165	jku	T2	R3	0.191	jku	T2	R1	0.264	jku	T2	R1	0.214
dbis	T1	R3	0.070	dbis	T1	R3	0.144	dbis	T1	R3	0.155	dbis	T1	R3	0.153
dbis	T1	R4	0.070	dbis	T1	R4	0.144	dbis	T1	R4	0.155	dbis	T1	R4	0.153
dbis	T1	R5	0.070	dbis	T1	R5	0.144	dbis	T1	R5	0.155	dbis	T1	R5	0.153
dbis	T2	R3	0.070	dbis	T2	R3	0.144	dbis	T2	R3	0.155	dbis	T2	R3	0.153
								dbis	T1	R2	0.094				
								dbis	T2	R2	0.094				
BETTER THAN BASELINES															
dbis	T1	R2	0.027	dbis	T1	R2	0.036	dbis	T2	R4	0.055	dbis	T1	R2	0.070
dbis	T2	R2	0.027	dbis	T2	R2	0.036	dbis	T2	R5	0.055	dbis	T2	R2	0.070
				icsi	T1	R3	0.035	dbis	T1	R1	0.044	dbis	T2	R5	0.059
				icsi	T1	R4	0.029	dbis	T2	R1	0.044	dbis	T2	R4	0.054
				dbis	T1	R1	0.029	icsi	T1	R3	0.041	icsi	T1	R3	0.049
				dbis	T2	R1	0.029	icsi	T1	R4	0.035	icsi	T1	R4	0.044
				dbis	T2	R5	0.026	icsi	T1	R5	0.031	dbis	T1	R1	0.041
				icsi	T1	R5	0.025					dbis	T2	R1	0.041
				dbis	T2	R4	0.022					icsi	T1	R5	0.037
SIMILAR TO BASELINES															
dbis	T2	R5	0.019	icsi	T1	R2	0.016	icsi	T1	R1	0.011	kart	T1	R3	0.019
dbis	T2	R4	0.017	icsi	T1	R1	0.016	icsi	T1	R2	0.009	icsi	T1	R1	0.019
dbis	T1	R1	0.010	baseline	T1	R1	0.010	kart	T1	R2	0.008	icsi	T1	R2	0.018
dbis	T2	R1	0.010	kart	T1	R2	0.008	kart	T1	R1	0.008	kart	T1	R1	0.016
baseline	T1	R1	0.006	kart	T1	R1	0.008	baseline	T1	R1	0.007	kart	T1	R4	0.014
kart	T1	R2	0.003	kart	T1	R3	0.007	kart	T1	R3	0.006	kart	T1	R2	0.013
kart	T1	R1	0.002	kart	T1	R4	0.006	kart	T1	R4	0.004	baseline	T1	R1	0.010
kart	T1	R3	0.002	samirit	T1	R1	0.005	samirit	T1	R1	0.004	samirit	T1	R1	0.005
samirit	T1	R1	0.002	baseline	T1	R2	0.002	baseline	T1	R2	0.002	baseline	T1	R2	0.002
kart	T1	R4	0.001												
baseline	T1	R2	0.001												

The 2017 task highlights

- The task is challenging!
- Subgenre recognition task is more difficult than genre
- High recall, but poor precision for many systems
- Systems should exploit hierarchies more
- No significant improvement from combining genre sources yet (Subtask 2)

AcousticBrainz audio commons upf. Universitat Pompeu Fabra Barcelona MTG Music Technology Group TUDelft tagtraum industries

https://multimediaeval.github.io/2017-AcousticBrainz-Genre-Task/