The Web Conference

Lyon April 27, 2018

Challenge Overview: Learning to Recognize Musical Genre from Audio

Michaël Defferrard

École Polytechnique Fédérale de Lausanne (EPFL)

Joint work with Sharada P. Mohanty (EPFL)
Sean F. Carroll (EPFL)
Marcel Salathé (EPFL)

Outline

FMA: a dataset for music analysis

Challenge design

Results

Fei-Fei Li
@drfeifei

Following

Datasets play crucial roles in advancing AI. #ImageNet helped to enable the latest deep learning advances.

Datasets Over Algorithms

Content without method leads to fantasy; method without content to empty sophistry. — Johann Wolfgang von Goethe ("Maxims and Reflections", 1892) "Perhaps the most important news of ...

spacemachine.net

5:12 AM - 22 Apr 2016

103 Retweets **174** Likes

2 103 174

Motivations

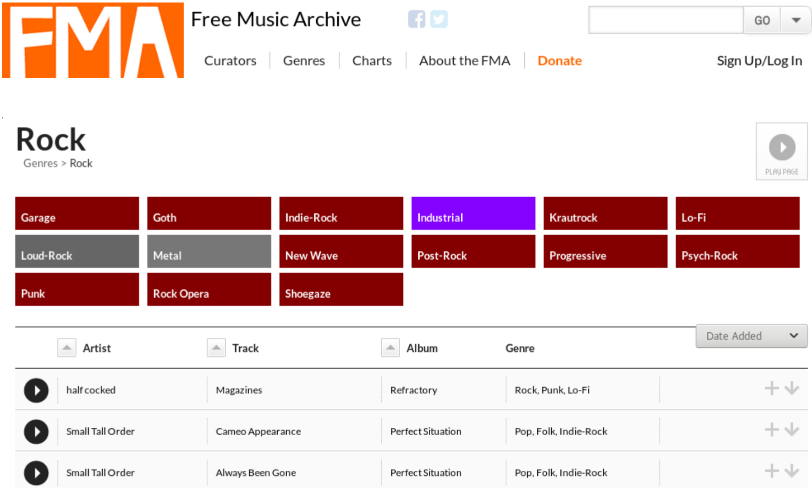
Goal: open dataset for Music Information Retrieval (MIR)
and Machine Learning (ML)

- Accelerate machine learning research on audio.
- Promote open data and open evaluation.
- Need both open benchmarking and standardized challenges.

There is a lot to learn from so much data!

The Free Music Archive

https://freemusicarchive.org



Data: key numbers & subsets

- 106,574 tracks from 16,341 artists and 14,854 albums.
- 917 GiB and 343 days of Creative Commons-licensed audio.
- Arranged in a hierarchical taxonomy of 161 genres.
- 518 pre-computed features per track.

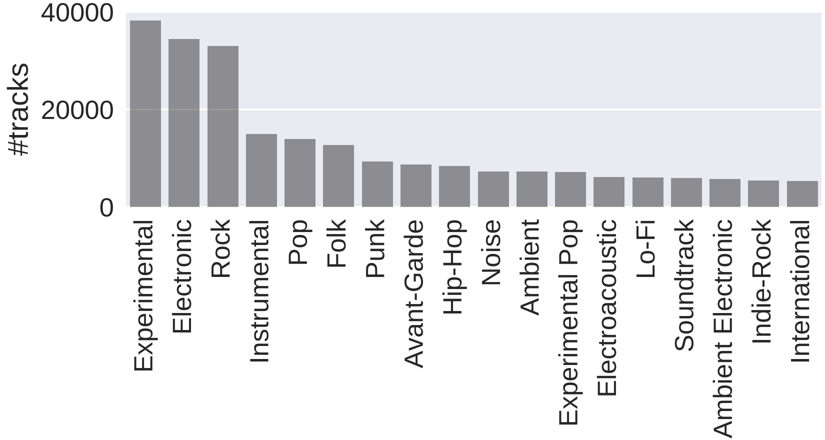
dataset	clips	genres	length	size	
			[s]	[GiB]	#days
small	8,000	8	30	7.4	2.8
medium	25,000	16	30	23	8.7
large	106,574	161	30	98	37
full	106,574	161	278	917	343

Data: metadata

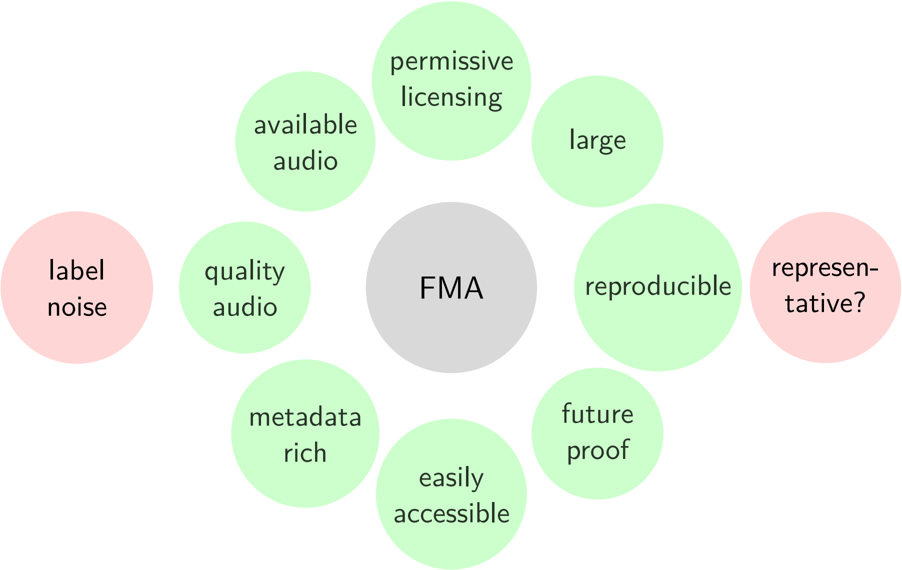
100% track_id	100% title	93% number
2% information	14% language_code	100% license
4% composer	1% publisher	1% lyricist
98% genres	98% genres_all	98% genres_top
100% duration	100% bit_rate	100% interest
100% #listens	2% #comments	61% #favorites
100% date_created	6% date_recorded	22% tags
100% album_id	100% title	
94% type	96% #tracks	
76% information	16% engineer	18% producer
97% #listens	12% #comments	38% #favorites
97% date_created	64% date_released	18% tags
100% artist_id	100% name	25% members
38% bio	5% associated_labels	
43% website	2% wikipedia_page	
	5% related_projects	
37% location	23% longitude	23% latitude
11% #comments	48% #favorites	10% tags
99% date_created	8% active_year_begin	
	2% active_year_end	

Data: genre distribution

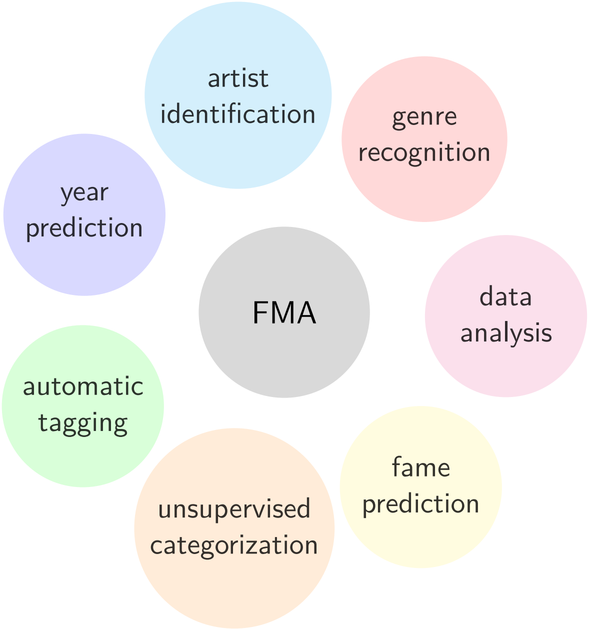
Large class imbalance!



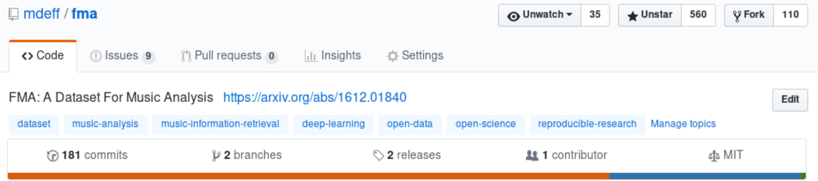
Qualities & Limitations

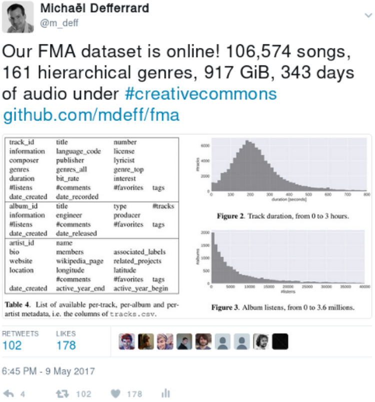


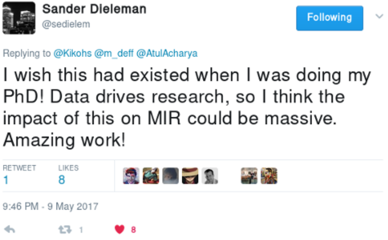
Usage

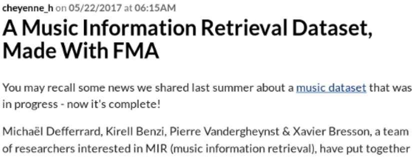

artist identification
genre recognition
year prediction
FMA
data analysis
automatic tagging
unsupervised categorization
fame prediction

People like it!









Outline

FMA: a dataset for music analysis

Challenge design

Results

Challenge design

Main issue: all the data is public

Solution: **two rounds**

- Round 1: public leaderboard, participants submit predictions
- Round 2: private test set, participants submit self-contained prediction systems

Challenge design: round 1

- Good to get continuous feedback.
- Used as a validation set to tweak the models.
- Cheating is possible by fingerprinting.
  Mitigated by providing many clips.

Challenge design: round 2

Real unseen test set $\rightarrow$ used to rank participants

- Detect overfitting (or cheating) on validation set.
- Round2 test is significantly harder than round1.
- No cheating as systems are run by ourselves.
- Remaining issue: participants might have trained on a larger dataset than provided.

Emphasis on openness and reproducibility

- Open data.
- Open evaluation.
- All systems are open source and fully reusable.
- The challenge is completely reproducible.

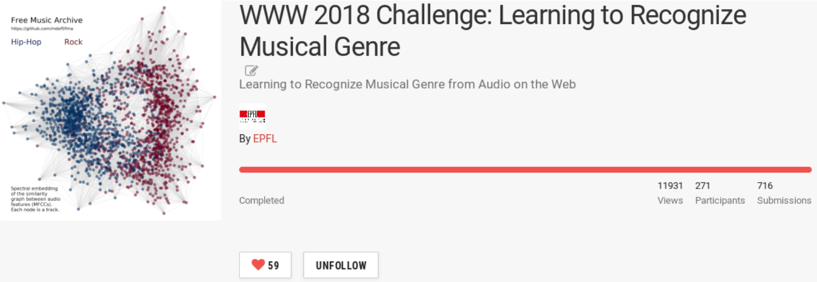
Outline

FMA: a dataset for music analysis

Challenge design

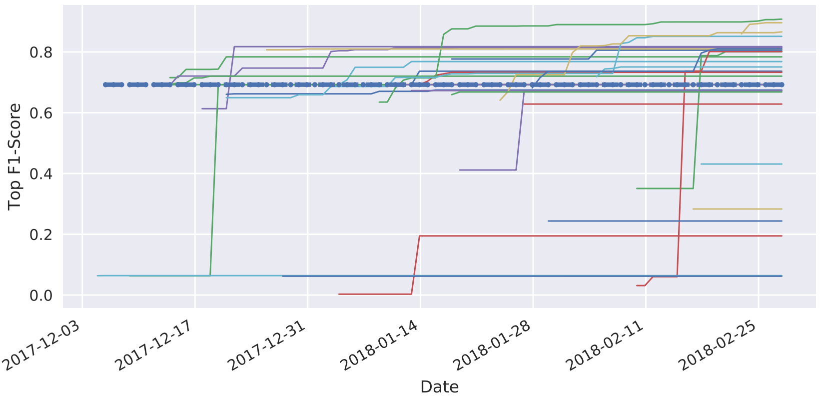
Results

Participation



- First round: 716 submissions from 38 participants
- Second round: 6 submitted systems

First round: public and continuous evaluation



First round: leaderboard

Δ	#	Participant	Log Loss	F1 Score	Entries	Last Submission (UTC)
	01.	iyjeong	0.33	0.909	54	Thu, 1 Mar 2018 08:36
	02.	hglim	0.333	0.923	110	Thu, 1 Mar 2018 12:45
	03.	mimbres	0.41	0.899	22	Thu, 1 Mar 2018 05:14
	04.	minzwon	0.55	0.851	47	Sat, 10 Feb 2018 22:43
	05.	ryantl	0.629	0.814	26	Wed, 24 Jan 2018 04:19
	06.	PKU_DL	0.639	0.817	12	Wed, 27 Dec 2017 12:45
	07.	AlgoHunt	0.659	0.8	34	Tue, 20 Feb 2018 07:25
	08.	Leukas	0.66	0.805	16	Fri, 9 Feb 2018 15:23
	09.	check	0.662	0.801	18	Thu, 1 Mar 2018 11:26
	10.	Jaehun	0.664	0.807	69	Mon, 25 Dec 2017 17:44
	11.	gg12	0.68	0.81	33	Sun, 18 Feb 2018 21:34
	12.	Philipp	0.82	0.751	17	Wed, 7 Feb 2018 16:29
	13.	Benjami...	0.824	0.742	30	Tue, 19 Dec 2017 07:55

First round: confusion matrix

True label \ Predicted label	Blues	Classical	Country	Easy Listening	Electronic	Experimental	Folk	Hip-Hop	Instrumental	International	Jazz	Old-Time / Historic	Pop	Rock	Soul-RnB	Spoken
Blues	4	0	0	0	0	1	0	0	0	0	0	0	0	2	0	0
Classical	0	66	0	0	1	3	2	0	1	1	0	0	0	1	0	0
Country	0	0	12	0	0	0	1	0	0	0	0	0	0	4	0	0
Easy Listening	0	0	0	2	2	1	0	0	0	1	0	0	0	0	0	0
Electronic	0	2	0	0	724	26	3	8	11	1	0	0	3	15	0	0
Experimental	0	2	0	0	26	211	7	2	4	1	0	0	1	15	0	0
Folk	0	1	0	0	3	9	142	0	2	1	0	1	2	17	0	0
Hip-Hop	0	0	0	0	26	5	1	232	0	0	0	0	0	4	0	2
Instrumental	0	2	0	0	18	11	4	0	116	0	0	0	1	14	0	0
International	0	1	0	0	8	4	3	2	1	94	0	0	0	6	0	1
Jazz	0	3	0	0	7	5	2	0	3	1	26	0	0	5	0	0
Old-Time / Historic	0	0	0	0	0	1	0	0	0	0	0	53	0	0	0	0
Pop	0	0	0	0	20	5	5	4	2	2	0	0	69	35	0	0
Rock	0	1	0	0	28	16	8	1	4	2	0	0	2	693	0	2
Soul-RnB	0	0	0	0	2	1	0	1	0	0	0	0	0	4	11	0
Spoken	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	14

Second round: submitted systems

1. Transfer Learning of Artist Group Factors to Musical Genre Classification by Jaehun Kim (@jaehun), TU Delft and Minz Won (@minzwon), Universitat Pompeu Fabra.
2. Ensemble of CNN-based Models using various Short-Term Input by Hyungui Lim (@hglim), `http://cochlear.ai`.
3. Detecting Music Genre Using Extreme Gradient Boosting by Benjamin Murauer, Universität Innsbruck.
4. ConvNet on STFT spectrograms by Daniyar Chumbalov (@check), EPFL and Philipp Pushnyakov (@gg12), Moscow Institute of Physics and Technologies (MIPT).
5. Xception[1] on mel-scaled spectrograms by @viper and @algohunt.
6. Audio Dual Path Networks[2] on mel-scaled spectrograms by Sungkyun Chang (@mimbres), Seoul National University.

[1] Xception: Deep Learning with Depthwise Separable Convolutions, `https://arxiv.org/pdf/1610.02357`

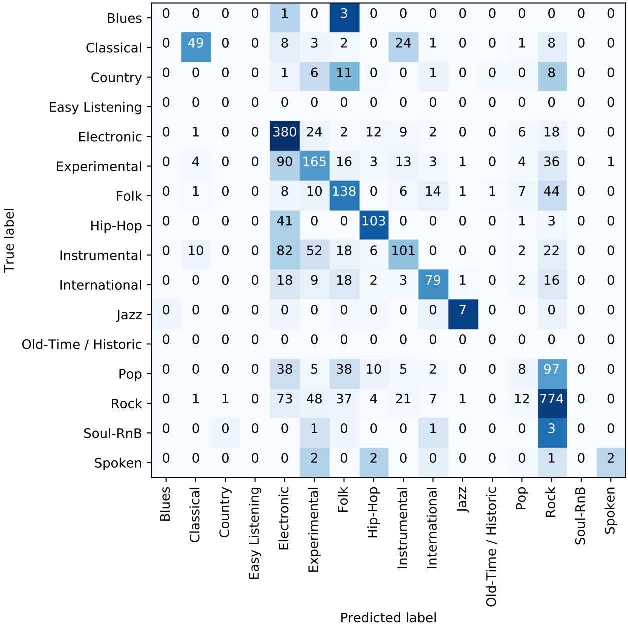
[2] Dual Path Networks, `https://arxiv.org/abs/1707.01629`

Second round: systems' performance

	Log loss				F1 score		
Team	R1a	R1b	R2	Rank	R1a	R1b	R2
minzwon & jaehun	0.55	0.67	**1.31**	1	85%	80%	63%
hglim	0.33	0.34	**1.34**	2	92%	92%	64%
benjamin_murauer	0.82	0.86	**1.44**	3	74%	74%	60%
gg12 & check	0.66	0.49	**1.50**	4	80%	86%	61%
viper & algohunt	0.66	0.65	**1.52**	5	80%	81%	60%
mimbres	0.41	0.43	**2.08**	6	90%	90%	60%

- *R1a* references the best scores obtained on the public leaderboard during the first round. The test data consisted of 35000 clips.
- *R1b* references the scores obtained by the submitted systems on a 3000 clips subset of the public test set used in the first round.
- *R2* references the scores obtained by the submitted systems on a 3000 clips private test set collected for the second round.

Second round: confusion matrix



Conclusion

- Most participants did not alter their system for the second round.
- Systems are often trained to overfit on validation data.
- Most systems used Deep Learning.
- Deep Learning on audio is hard: much larger samples[3], not a lot of battle tested architectures.
- All participants used DL architectures on spectrograms, which are images.

[3] At a sampling rate of 22050 Hz, a 3 minutes song is ≈ 4 million floating point numbers. To be compared with $256 \times 256 \approx 65000$ pixels on ImageNet.

Lessons learned

- Challenge design is hard.
- There is a real interest in the ML community and it's fun.
- Organizing a challenge will take more time than you think.
- Challenges inspire young people and encourage them to learn.

Special thanks to all the participants!

Slides https://doi.org/10.5281/zenodo.1243501

Paper Defferrard, Benzi, Vandergheynst, Bresson,
FMA: A Dataset For Music Analysis, ISMIR, 2017
https://arxiv.org/abs/1612.01840

Paper Defferrard, Mohanty, Carroll, Salathé,
Learning to Recognize Musical Genre from Audio, WWW, 2018
https://arxiv.org/abs/1803.05337

Code & Data https://github.com/mdeff/fma

Challenge https://www.crowdai.org/challenges/www-2018-challenge-learning-to-recognize-musical-genre

https://github.com/crowdAI/crowdai-musical-genre-recognition-starter-kit

Thanks Questions?