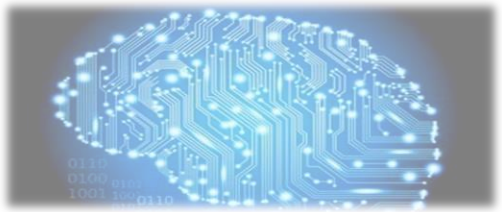
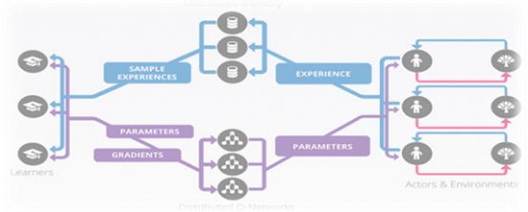

Musical Genre Classification via CNN

AutoML 연구실
석사 과정 조성현
21.10.19

The standard statistics of the datasets

Characteristics of datasets

Dataset	|W|	Avg. Length	Domain	|C|	Avg. Size of Class
Ballroom	698	30	Genre	8	87.25 ± 17.56
Ballroom-Extended	4180	30	Genre	13	321.54 ± 195.70
FMA-MEDIUM	24984	30	Genre	16	1561.50 ± 2069.36
FMA-SMALL	7996	30	Genre	8	999.50 ± 0.50
GTZAN	1000	30	Genre	10	100.000 ± 0.000
ISMIR04	2187	120	Genre	6	121.17 ± 95.60
MICM	1137	360	Genre	7	162.43 ± 119.72
Tropical Genres	1500	30	Genre	5	300.000 ± 0.000

|W| : the number of patterns
Avg. Length : the average length of music excerpts in the collection
Domain : the domain used for creating the corresponding dataset
|C| : the number of classes or tags
Avg. Size of Class : the average number of patterns per class and its standard deviation

Experiment Settings

- Overall preprocessings and settings
  1. Convert to .wav files
  2. Compute Mel Spectrograms and MFCCs and save as .npy files
     ($\mathrm{Length} \leq 30s$: zero padding, $\mathrm{Length} \geq 30s$: cut to $30s$)
  3. Cut into five splits to make five channels
     (0~10s, 5~15s, 10~20s, 15~25s, 20~30s)
  4. Depthwise convolution
  5. Multimodal CNN (Train : Test = 8 : 2)

Dataset	Avg. Length
Ballroom	30
Ballroom-Extended	30
FMA-MEDIUM	30
FMA-SMALL	30
GTZAN	30
ISMIR04	120
MICM	360
Tropical Genres	30

Comparison with other CNN models

Comparison of the result using Mel Spectrogram

Dataset	Proposed	ResNet18	ResNet50	GoogleNet	DenseNet	ShuffleNet v2
Ballroom	-	-	-	-	-	-
Ballroom-Extended	-	-	-	-	-	-
FMA-MEDIUM	-	-	-	-	-	-
FMA-SMALL	-	-	-	-	-	-
GTZAN	64.0	67.0	65.5	57.5	70.5	62.5
ISMIR04	-	-	-	-	-	-
MICM	-	-	-	-	-	-
Tropical Genres	-	-	-	-	-	-

Comparison of the result using Mel Spectrogram and MFCC

Dataset	Proposed	DenseNet	MobileNet	ResNet	ShuffleNet-v2
Ballroom	-	-	-	-	-
Ballroom-Extended	-	-	-	-	-
FMA-MEDIUM	-	-	-	-	-
FMA-SMALL	-	-	-	-	-
GTZAN	86.4	74.3	63.6	70.4	63.8
ISMIR04	-	-	-	-	-
MICM	-	-	-	-	-
Tropical Genres	-	-	-	-	-