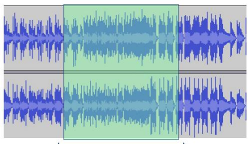
Music Popularity Prediction via Techniques in Deep Supervised Learning

Brian Rossi (brossi@stanford.edu); Mitchell Pleus (apleus@stanford.edu)
Stanford University

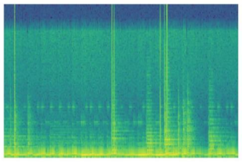
Stanford University

Overview

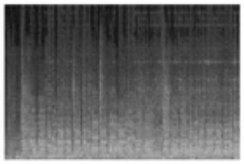
The problem we investigated for this project is what makes a song popular? More specifically, can we discern from a raw audio signal and its underlying characteristics whether or not a song will be popular? Based on prior research in genre classification [4][5], our strategy was to preprocess audio files by converting them into spectrogram images, and use a deep CNN as a multi-class classifier to predict the audio files' popularity via their associated spectrogram images. We used Google Cloud Engine to house our data, preprocess, and train our model remotely.



10 second mp3 clip



RGB spectrogram



Grayscale spectrogram

Class Labels (% of dataset)

High popularity (36.5%)

Medium popularity (46.7%)

Low popularity: (16.8%)

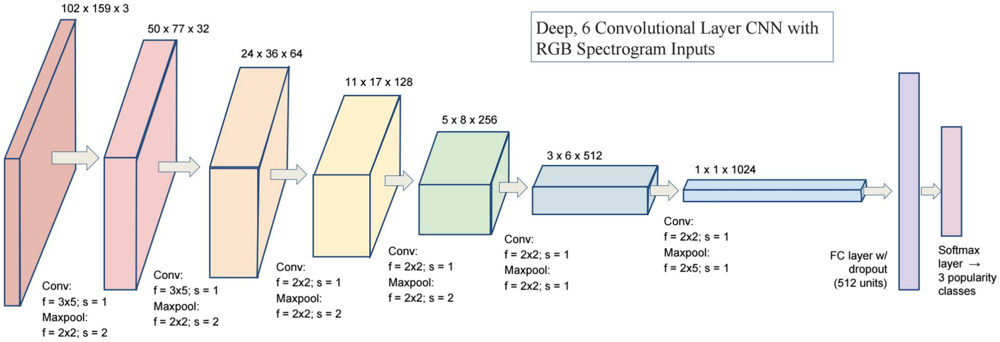
Data

The dataset we used was "FMA: A Dataset for Musical Analysis"[2] which consisted of 8,000 song mp3 files and associated popularity metadata.

Our strategy was to preprocess audio files by converting clips into RGB and grayscale spectrogram images for our deep CNN.

Popularity was represented as #listens in the metadata. Setting reasonable thresholds, we converted this into three popularity buckets.

Best Model Architecture



Results

Hyperparameters for Best Model		
Learning Rate	**Batch Size**	**Dropout Rate**
5 e-4	16	0.4

Accuracies			
Architecture	**Training**	**Validation**	**Test**
2-Conv Layer CNN (mono)	44%	41%	42%
4-Conv Layer CNN (mono)	46%	51%	51%
4-Conv Layer CNN (RGB)	53%	55%	53%
6-Conv Layer CNN (RGB)	59%	63%	61%

We ran 4 different models for 100 epochs and tuned hyperparameters for each model individually. We split our 8,000 example dataset into 80% train, 10% validation, and 10% test.

Discussion and Future Work

From this project, we see that it is very difficult to predict the popularity of a song from the raw audio signal. We battled against underfitting, which is why in our later models we moved to deeper networks and transitioned from mono input to RGB input.

In future work, it would be interesting to see if incorporating more metadata (e.g. artist popularity, genre, etc.) would increase the accuracy in predicting music popularity.

Spectrograms can theoretically glean harmonic and rhythmic information, but more in depth analysis of different musical touchstones and characteristics could supplement future popularity predictors (e.g. NLP analysis of lyrics).

References

[1] L. Yang , S. Chou , J. Liu , Y. Yang , Y.Chen, Revisiting the Problem of Audio-Based Hit Song Prediction Using Convolutional Neural Networks, 2017.

[2] M. Defferrard, K.Benzi, P. Vandergheynst, X. Bresson, FMA: A Dataset for Music Analysis, 2017.

[3] J. Pham, E. Kyauk, E. Park, Predicting Song Popularity, 2015.

[4] M. Meza, J. Nalianya, Music Genre Classification Using Deep Learning, 2018.

[5] A O'Beirne, A. Zamora, Music Genre Classification Using Mel Spectrogram Representations, 2018.