Project Report on
Music Genre Classification

Kunal Sah Ritam Chattopadhyay

1. ABSTRACT

This project is based on Music Genre Classification using various machine learning and deep learning techniques. We follow two different techniques to classify the genres of the music files, namely the K-Nearest-Neighbours technique and the Convolutional Neural Networks. We implement these two widely followed techniques for our problem statement and compare their outcomes.

2. TECHNICAL DETAILS

Dataset used is GTZAN, which includes music form 10 different genres, each having 100 .wav files. Different genres are: "hip-hop","disco","country","blues","jazz","classical","pop", "reggae","rock","metal".

2a. Spectrogram and Mel Spectrogram

The Short-Time Fourier Transform (STFT) of the signal is taken within fft=2048 and hop size=512 and a Hann window. Next, we compute the power spectrum and then apply the triangular MEL filter bank. (Bahuleyan, 2018)This is followed by taking the discrete cosine transform of the logarithm of all filter bank energies, thereby obtaining the MFCCs. The parameter n_mels, which corresponds to the number of filter banks, was set to 128 in this project.

2b. Implementation of KNN

The Mel Spectrograms are of dimension 1293×128. For computing the distance between Mel Spectrograms, we used the KLdivergence as the metric. The KLdivergence is computed by using the multivariate Normal Distribution pdf with means of the Spectrograms $\mu_1, \mu_2 \in R^{128}$ and Covariance matrices $\Sigma_1, \Sigma_2 \in R^{128 \times 128}$.

$D_{KL}(\mathcal{N}_1||\mathcal{N}_2) = \frac{1}{2}(tr(\Sigma_1^{-1}\Sigma_2) + (\mu_1 - \mu_2)^T\Sigma_1^{-1}(\mu_1 - \mu_2) - k + \log(\frac{|\Sigma_1|}{|\Sigma_2|}))$.(Wikipedia contributors, 2021)

Then following the standard KNN algorithm, with K = 3.

2c. Implementation of CNN

Convolution layer 1: Kernel = 32,Kernel size= (3,3) padding = 1, Max Pooling = (4,4)

Convolution layer 2: Kernel = 64, Kernel size = (3,3) padding = 1, Max Pooling = (2,2)

Convolution layer 3: Kernel = 128, Kernel size = (3,3), padding = 1, Max Pooling = (2,2)

Feed-Forward Neural Network Architecture: I/P layer = 71680, Hidden Layers = [1000,500,100], O/P Layer = 10, with dropout = 20 %. Learning rate = 0.001.

3. RESULTS

Training Set = 75% of Data set, Test Set = 25% of Data Set.

Results of CNN implementation:

Training Loss plot with Epoch:

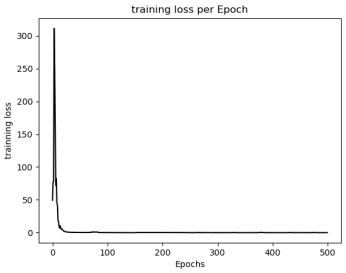

Table for Model Summary after training with CNN

Class	Precision	Recall	F1-score
0	0.89	0.65	0.76
1	0.72	0.54	0.62
2	0.83	0.66	0.73
3	0.52	0.88	0.66
4	0.63	0.75	0.69
5	0.81	0.50	0.62
6	0.86	0.70	0.78
7	0.91	0.80	0.85
8	0.68	0.85	0.75
9	0.29	0.50	0.36

CNN Accuracy: **70%**

Comparison of Confusion matrices of KNN and CNN trained model on test set

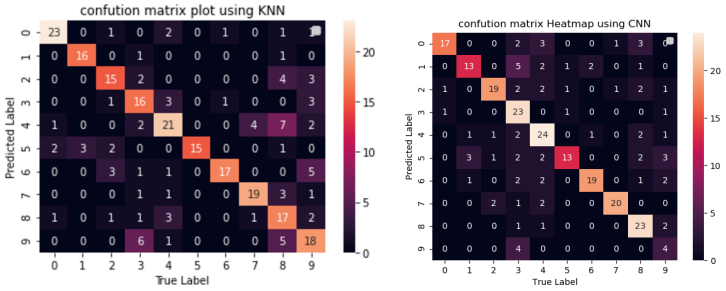

Table for Model Summary after training with KNN

Class	Precision	Recall	F1-score
0	0.85	0.79	0.82
1	0.84	0.89	0.86
2	0.65	0.62	0.64
3	0.53	0.67	0.59
4	0.66	0.57	0.61
5	1.00	0.65	0.79
6	0.89	0.63	0.74
7	0.79	0.76	0.78
8	0.44	0.65	0.52
9	0.51	0.60	0.55

K-Nearest-Neighbors (KNN) Accuracy: **67%**

4. NOVEL CONTRIBUTIONS

The paper we followed only gives CNN as a solution. In extra we also Implemented k-nearest neighbors (KNN), which can also be a solution to classify music genre as the test accuracy of KNN is close to CNN. Here we assumed each Mel Spectrogram belongs to a Multivariate Gaussian Distribution with a mean of 13 dimension and a covariance of 13X13 dimension. For calculating distance between to Gaussian distribution used KL divergence. And showed comparison between KNN and CNN.

Individual Contribution:

1. Kunal Sah

Implemented Convolutional Neural Network to classify music genres using Pytorch .

2. Ritam Chattopadhay

Implemented KNN model to classify music genres using Numpy.

5. TOOLS USED

1.Spectrogram and MFCC computation:librosa,scipy.io.wav, python_speech_features.
2. KNN Algorithm: Python, Numpy.
3. CNN implementation: Pytorch, Numpy.
4. Additional libraries used: pickle,os.
5. Plots: matplotlib,seaborn.

References

Bahuleyan, H. Music genre classification using machine learning techniques. *arXiv preprint arXiv:1804.01149*, 2018.

Wikipedia contributors. Kullback–leibler divergence — Wikipedia, the free encyclopedia, 2021. URL https://en.wikipedia.org/w/index.php?title=Kullback%E2%80%93Leibler_divergence&oldid=1029607529. [Online; accessed 21-June-2021].