https://doi.org/10.36375/prepare_u.a65

IMPLEMENTATION OF MACHINE LEARNING AND DEEP LEARNING IN IMPROVING SIGNAL TO NOISE RATIO FOR AUDIO ANALYSIS IN DIGITAL MEDIA PLATFORMS

Sudipta Ghosh, Subhojit Jalal, Rajat Agarwal, Kalyan Chatterjee, Sayanti Banerjee
Department of Mechanical and Automation Engineering
Department of Electronics and Communication Engineering
Amity University Kolkata
Major Arterial Road, Action Area11, New Town, Kolkata 700135
sudiptaghosh1510@gmail.com

Abstract-Audio analysis is fast becoming a requirement of digital media for analysing multiple frequencies of sound at a time and also reduce the background noise.The urban sound challenge is increasing day by day and the problem is meant to introduce for audio analysis and processing in the usual classification scenario. Model is implemented on the basis of keras framework and librosalibrary. Keras is capable of running of the algorithm tensorflow. Kerascan be described as an interface rather than a standalone machine learning tools. Librosa is one of the python library for music and audio analysis. It helps us with necessary music information retrieval systems.

The authorcollect the database and use the data and also use the graph for a better understanding of audio data analysis.

KEYWORDS: os; pandas; librosa; keras; glob; numpy; sklearn

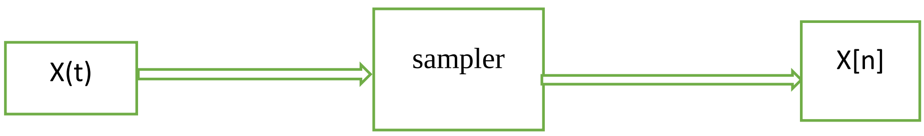
INTRODUCTION:

Our sense of hearing gives us the information of our environment with respect to the characteristics of sound producing objects present in our locality. We can easily hear the twittering of birds outside and also the movement of traffic at a distance while doing our own work. But it is not easy to imagine that the machine also understand, hear and also make the difference of sound and their multiple frequencies at a time. There are many important techniques present in audio signal processing – audio data compression, synthesis and audio classification.

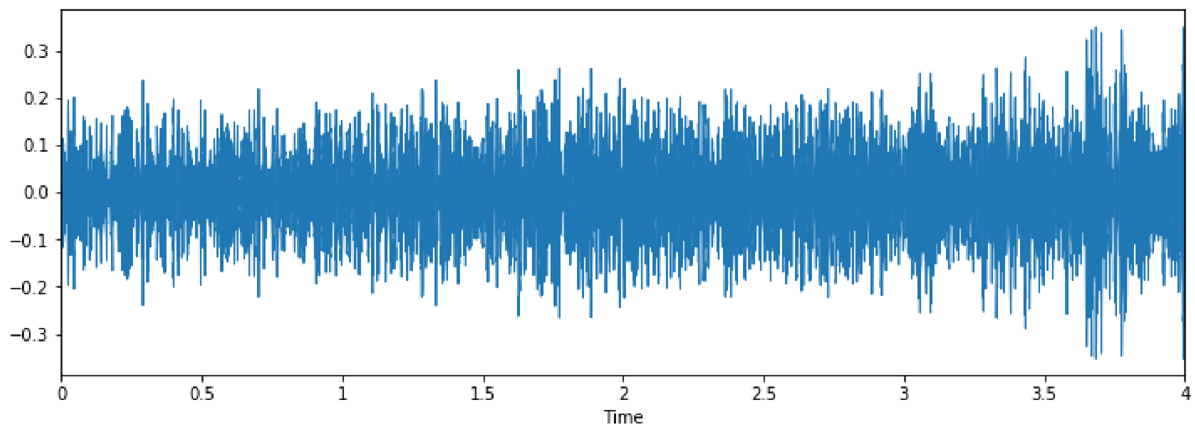
Audio analysis is the extraction of information and meaning from audio signals for analysis, synthesis and classification, storage. The observation mediums and the numerical methods vary as audio analysis can refer to the human ear and how people interpret with the audible sound sources. Once theinformation of audio signal has been observed, only then it can be proceed for the logical, emotional, descriptive or otherwise relevant interpretation by the authors.

As we are directly and indirectly connected with the audio. as our brain works continuously with the different frequencies of sounds. Thus, our brain starts to understands the information.

Here we also use the sampling method for audio analysis. Sampling is the process by which a continuous time signal can be converted into discrete time signal.



https://doi.org/10.36375/prepare_u.a65



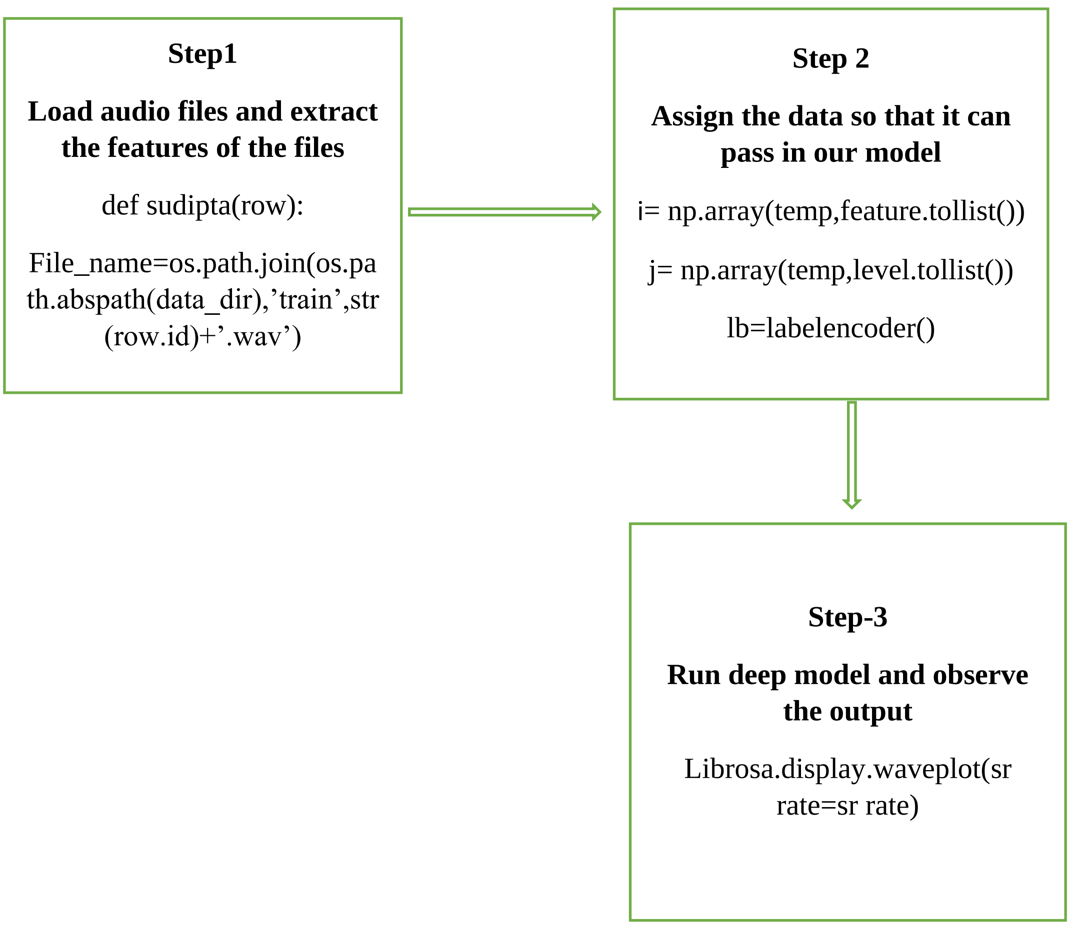
Sampling rate

Here the authors extract the values of audio file at half a second. It is called sampling of audio data and the rate at which the audio data is sampled is called sampling rate.

On the contrary, here the author represents the data in frequency domain.

Applications-1) indexing of many music collection according to their audio feature2) Speech recognition, processing and synthesis.

Flow chart

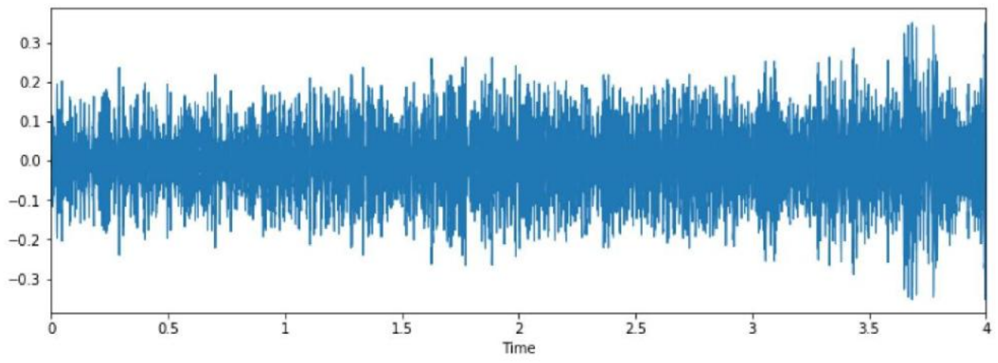


https://doi.org/10.36375/prepare_u.a65

Output-

<u>Output 1</u>

```
Out[4]: <matplotlib.collections.PolyCollection at 0x1f58631dda0>
```



<u>output 2</u>

	ID	Class
0	0	siren
1	1	street_music
2	2	drilling
3	3	siren
4	4	dog_bark
5	6	children_playing
6	10	street_music
7	11	drilling
8	12	gun_shot
9	15	dog_bark
10	17	engine_idling
11	18	siren
12	19	gun_shot
13	20	engine_idling
14	22	air_conditioner
15	24	air_conditioner
16	26	engine_idling
17	27	siren
18	32	children_playing
19	33	jackhammer

https://doi.org/10.36375/prepare_u.a65

```
20      35         gun_shot
21      36            siren
22      37       jackhammer
23      38     street_music
24      40       jackhammer
25      42     engine_idling
26      43            siren
27      44  children_playing
28      45       jackhammer
29      46     engine_idling
...    ...              ...
5405  8691     street_music
5406  8693         car_horn
5407  8694     street_music
5408  8695  air_conditioner
5409  8699     street_music
5410  8701  air_conditioner
5411  8703         gun_shot
5412  8704            siren
5413  8705       jackhammer
5414  8706            siren
5415  8707         gun_shot
5416  8709         dog_bark
5417  8710     street_music
5418  8711         dog_bark
5419  8712       jackhammer
5420  8713         dog_bark
5421  8714     engine_idling
5422  8715            siren
5423  8716  children_playing
5424  8717       jackhammer
5425  8720     street_music
5426  8721     engine_idling
5427  8722         dog_bark
5428  8723     street_music
5429  8724  children_playing
5430  8725     engine_idling
5431  8726         dog_bark
5432  8727     engine_idling
5433  8728     engine_idling
5434  8729  air_conditioner

[5435 rows x 2 columns]
```

<u>Output 3</u>

```
[[ 0.  0.  1. ...,  0.  0.  0.]
 [ 0.  0.  0. ...,  1.  0.  0.]
 [ 0.  0.  0. ...,  0.  0.  0.]
 ...,
 [ 0.  1.  0. ...,  0.  0.  0.]
 [ 0.  1.  0. ...,  0.  0.  0.]
 [ 0.  0.  0. ...,  0.  0.  0.]]
```

Output 4

```
Epoch 21/50
5435/5435 [==============================] - 1s 176us/step - loss: 0.8294 - acc: 0.7165
Epoch 22/50
5435/5435 [==============================] - 1s 178us/step - loss: 0.7960 - acc: 0.7323
Epoch 23/50
5435/5435 [==============================] - 1s 177us/step - loss: 0.7805 - acc: 0.7420
Epoch 24/50
5435/5435 [==============================] - 1s 175us/step - loss: 0.7629 - acc: 0.7426
Epoch 25/50
5435/5435 [==============================] - 1s 177us/step - loss: 0.7702 - acc: 0.7387
Epoch 26/50
5435/5435 [==============================] - 1s 176us/step - loss: 0.7355 - acc: 0.7433
Epoch 27/50
5435/5435 [==============================] - 1s 175us/step - loss: 0.7291 - acc: 0.7501
Epoch 28/50
5435/5435 [==============================] - 1s 175us/step - loss: 0.7226 - acc: 0.7533
Epoch 29/50
5435/5435 [==============================] - 1s 176us/step - loss: 0.7046 - acc: 0.7592
Epoch 30/50
5435/5435 [==============================] - 1s 173us/step - loss: 0.6785 - acc: 0.7650
Epoch 31/50
5435/5435 [==============================] - 1s 174us/step - loss: 0.6750 - acc: 0.7674
Epoch 32/50
5435/5435 [==============================] - 1s 175us/step - loss: 0.6738 - acc: 0.7711
Epoch 33/50
5435/5435 [==============================] - 1s 176us/step - loss: 0.6368 - acc: 0.7890
Epoch 34/50
5435/5435 [==============================] - 1s 174us/step - loss: 0.6557 - acc: 0.7798
Epoch 35/50
5435/5435 [==============================] - 1s 172us/step - loss: 0.6427 - acc: 0.7847
Epoch 36/50
5435/5435 [==============================] - 1s 174us/step - loss: 0.6295 - acc: 0.7910
Epoch 37/50
5435/5435 [==============================] - 1s 174us/step - loss: 0.6016 - acc: 0.7941
Epoch 38/50
5435/5435 [==============================] - 1s 174us/step - loss: 0.6001 - acc: 0.7906
Epoch 39/50
5435/5435 [==============================] - 1s 173us/step - loss: 0.5932 - acc: 0.8004
Epoch 40/50
5435/5435 [==============================] - 1s 173us/step - loss: 0.5901 - acc: 0.8055
Epoch 41/50
5435/5435 [==============================] - 1s 173us/step - loss: 0.5963 - acc: 0.7950
Epoch 42/50
5435/5435 [==============================] - 1s 172us/step - loss: 0.5979 - acc: 0.7991
Epoch 43/50
5435/5435 [==============================] - 1s 174us/step - loss: 0.5686 - acc: 0.8035
Epoch 44/50
5435/5435 [==============================] - 1s 172us/step - loss: 0.5868 - acc: 0.8007
Epoch 45/50
5435/5435 [==============================] - 1s 173us/step - loss: 0.5697 - acc: 0.8070
Epoch 46/50
5435/5435 [==============================] - 1s 172us/step - loss: 0.5567 - acc: 0.8099
Epoch 47/50
5435/5435 [==============================] - 1s 173us/step - loss: 0.5709 - acc: 0.8072
Epoch 48/50
5435/5435 [==============================] - 1s 170us/step - loss: 0.5517 - acc: 0.8098
Epoch 49/50
5435/5435 [==============================] - 1s 173us/step - loss: 0.5529 - acc: 0.8072
```

Conclusion-

Audio analysis has been developed steadily over the last few years and it is incorporated into several new applications. In this paper, authors have developed the audio analysis and it's required graph using machine learning and deep learning tools. The dataset contains 8732 different sounds <=4s.

Reference-

[1] Gideon mendels,"how to apply deep learning and machine learning method in audio analysis", Nov,2019

[2] Acton, Ciaran," analysis of variance(ANOVA)",2009

[3] Mitchell,"machine learning",1997

[4] Bishop,C.M,"pattern recognition and machine learning",2006

[5] Schmidhuber, “deep learning in neural network”,2015

[6] FaizanShaikh,”getting started with audio data analysis using deep learning”, August,2017

[7] G.W.Smith,” The machine as Artist”,April,2017

[8] Shen,yelong,”a latent semantic model with convolution-pooling structure for information retrieval”,October,2013

[9] Jotikasingh,”introduction to audio processing”,September,2019

[10] A.Huang,andR.Wu,”deep learning for music”

[11] N.Jaitly and G.Hinton,”learning a better representation og speech soundwaves using neural networks”,2011