International Journal for Innovative Engineering and Management Research

PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL

www.ijiemr.org

IJIEMR Transactions, online available on 8th Aug 2024. Link

https://ijiemr.org/downloads.php?vol=Volume-13& issue=issue08

DOI: 10.48047/IJIEMR/V13/ISSUE 08/15

Title Audio Annotation and Classification of Tollywood Music Using Deep Learning

Volume 13, ISSUE 08, Pages: 113 - 118

Paper Authors
Dr. G.V. Ramesh Babu , P. Keerthi

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Audio Annotation and Classification of Tollywood Music Using Deep Learning

Dr. G.V. Ramesh Babu
Associate Professor, Department of Computer Science, Sri Venkateswara University, Tirupati
gvrameshbabu74@gmail.com
P Keerthi
Master of Computer Applications,
Sri Venkateswara University, Tirupati
keerthiparri19@gmail.com

Abstract
The project aims to classify Telugu songs into vocal and non-vocal compo- nentsusing machine learning. Initially, a dataset of thirty songs is prepared with ten songs each from different genres. Each song is manually labeled for vocal and non-vocal regions using an open source tool called wavesurfer. The training set was used to train the machine learning model. The ma- chine learning model explored was Artificial Neural Network (ANN) to classify the vocals and non- vocals in a song. The classification further helps in Instrument identification, singer/ vocalist identification and to develop search engines.

Introduction

Entertainment plays a major role in human life as it holds the attention and interest of an audience. Movies, sports, music and hobbies are some of the means of entertainment. Music is a form of art and a cultural activity which impact the society culturally, morally and emotionally. In the field of music there are wide varieties of genres ranging from traditional music to pop music. Music has several components like instruments and vocals. The parts of the song which are sung by a vocalist with or without accompanying instruments are referred as vocals. And the parts with just instrumental music is called non- vocals. In the field of music there are wide varieties of genres ranging from traditional music to pop music. The popular genres in kannada music and telugu music are pop, devotional, rock, sad, patriotic, bhavageete, melody, electronic, classical. In the recent days, there is a huge need for music information retrieval for instrument identification, singer/ vocalist identification and to develop search engines.Thus a learning model is developed to classify the vocal and non- vocal regions of the song

Objective

The aim of the project Vocal and non-vocal classification is to develop a machine learning model to classify the vocal and non-vocal regions of the audio file.

- Collect audio files
- Annotating the audio files as vocals and non- vocal regions
- Explore neural network models and develop a learning model

Tools and Technology

MATLAB (matrix laboratory) is a multi-paradigm numerical computing environment and proprietary programming language developed by Math- Works. MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and inter- facing with programs written in other languages, including C, C++, C#, Java Fortran and Python. MATLAB also provides signal processing tool- box which provides functions and apps to analyse, pre-process, and extract features from the 1D or 2D signals. Audio and video annotations are done using signal processing toolbox. To train a machine, we need to collect the data and label using tools. For our data we have used three different tools WaveSurfer & Sonic Visualiser for audio annotations and ImageLabeler tool for video annotations.

- Wavesurfer

WaveSurfer is an audio editor widely used for studies of waveform, spec- trum, pitch contour, transcriptions and many more features of audio signal in an interactive program which provides display of the sound signals. Fig- ure 1 showing the different panes that can be used in this tool for audio annotation by visualising the signal.

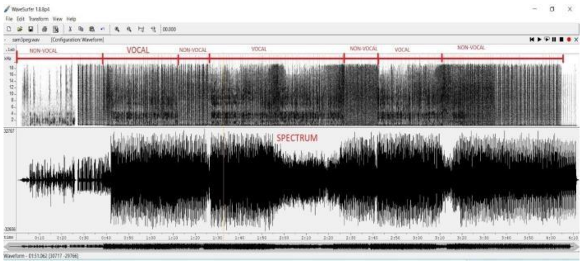

Figure 1: WaveSurfer tool

- Sonic Visualiser

Sonic Visualiser is also one of the audio editor tools used to visualise, anal- yse and used to annotate audio files or sound files with advanced technology and colourful display of the spectrum and waveforms of audio. Figure 2 showing the spectrum of the audio file.

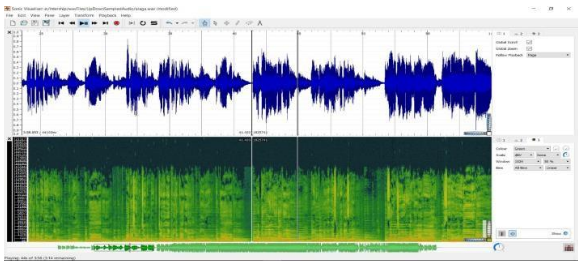

Figure 2: Sonic Visualiser tool

Both tools are used for the audio annotations in an easy way. We used WaveSurfer for annotation and saving transcription of audio files. Sonic Visualiser is used to verify the labelled audio files by visualising the spectrum and waveforms in coloured display. This helped us to reduce the error in transcribing (labelling) the audio files by adding some more labels after visualisation. Labelling of audio files is explained in chapter 4.

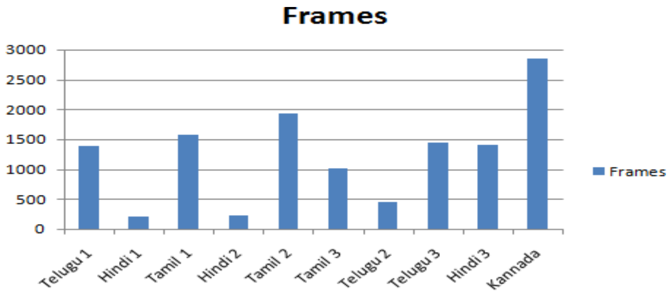

Figure 3: Chart representing number of frames in each Dataset

Organizing the Data

DATASET USED

As we are monitored to the Tollywood music (Telugu) the genres that we encounter in the telugu music that are present after the year of 2000 are,

- Classical:- It includes both the liturgical as well as the secular music.
- Electronic:- It includes mostly the electrical instruments and digital instruments.
- Pop:- It is also known as the popular music and it includes many different styles.
- Melody: - A melody is a combination of pitch and rhythm. Melodies often consist of one or more musical phrases or motifs, and are usually repeated throughout a composition in various forms.
- Sadcore :- It is a subgenre occasionally identified by music journalists to describe ex- amples of alternative rock characterised by bleak lyrics, downbeat melodies and slower tempos, or alternatively, songs with deceivingly upbeat melodies that are simultaneously characterised by depressive lyrical undertones or imagery.
- Rock:- Rock music is a broad genre of popular music that originated as ”rock and roll” and later developed into many different styles.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Collection of Data

The songs that will fall under the above mentioned genres are collected in such a way that each genre gets 10 songs respectively. And all the songs are taken that are composed after the year of 2000. The downloaded songs will be in the (.mp3) format which is the most commonly used format of the audio files in the present days so for the better experiment purpose we convert the (.mp3) file into the (.wav) format which is the most efficient and the actual format of the audio files which will have more clarity and less noise when compared to the (.mp3) files.

Obtaining other Files

As we are dealing with audio annotation part it is necessary for us to have the transcription file i.e; the file which contains the labellings of the particular audio file which is used in the further process of our experiment or task. The process of obtaining the transcriptions file (.lab) is explained in detailed later.

Data Statistics

The average Non vocals and vocals for a given genre i.e; in the genres that we have considered are represented as the bar graphs below which illustrates the clear count of the vocals and non vocals in the form of seconds that are present in the whole genre which contains 10 songs each. The next figure illustrates the songs that are taken under each genre are composed in which year and the number of songs that year contains

Table: the table consist of duration of vocal and non-vocal of each genre.

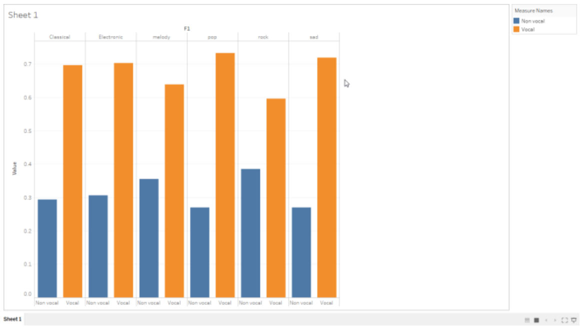

Figure 4: Here the y-axis represents the seconds and the x-axis represents both the genres and the Non vocal and vocal fields

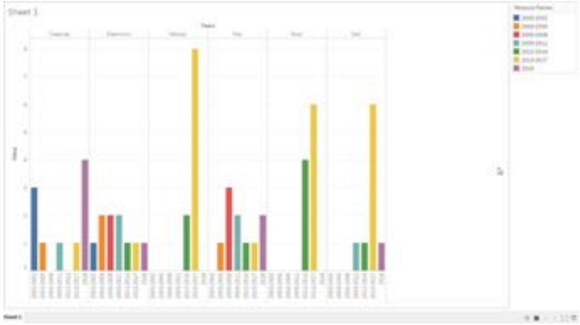

Figure 5: The above mentioned statistics are only for the telugu songs that have been taken.

➢ Labelling of Music Files

Vocal and Non-vocal regions of music files need to be labeled manually for supervised machine learning. Wavesurfer tool is used for this purpose. WaveSurfer is a popular tool used for audio editing, transcription and acoustic studies. It is simple to use but provides several functionalities It can display sound pressure waveforms, spectral sections, spectrograms, pitch tracks and transcriptions. It can be used to transcript the audio files in numerous formats. The music files are labeled using this tool. Labels are then stored in .lab files. The labeling scheme is as follows:

Table 1: Duration of Vocal and Non-Vocal for Audio Files.

Region	Melody	Rock	Sad	Classical	Electronic	Pop
Vocal	24:11	22:00	24:33	27:18	32:43	31:26
Non-Vocal	13:43	13:25	9:38	11:30	14:15	11:36

0: For non-vocals, i.e. all region except vocals. It contains region where only musical instrument is played.

1: For vocals, i.e. region where a singer is singing

The song labeling is done manually by listening. For our aid, Wavesurfer allows several options for better analysis of audio file. We can visualize au- dio in the form of waveform and spectrogram. Visualization of spectrogram aids in labeling of audio file while listening to music. We can zoom into spectrogram to label vocal and non vocal sections of music accurately.

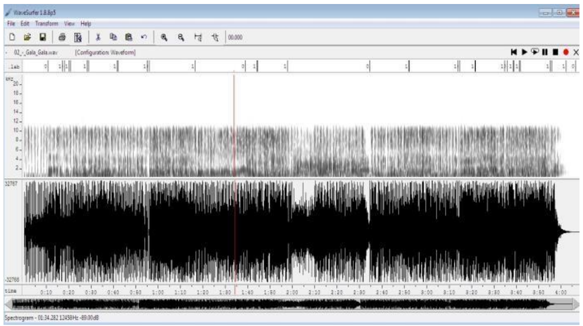

Figure 6: spectrogram of complete music file

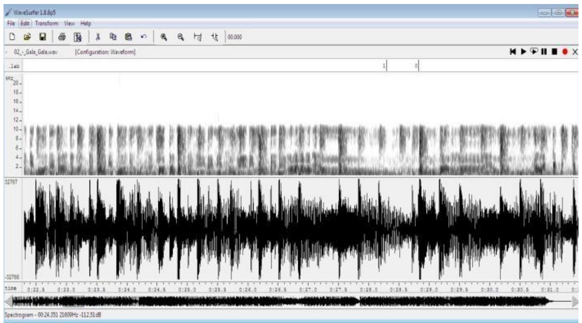

Figure 7: Zoomed section of spectrogram of a music file

Deep Learning
To extract the features and to train the data of audio file to classify them as vocal and non-vocal we use the concepts of deep learning.

Deep Learning
It is a machine learning technique that guides the system to do the tasks naturally as humans can do, learn by example. Driverless cars are built by using Deep learning methods, enabling them to recognize a stop sign, or to distinguish a pedestrian from a lamppost. It is the key to voice control in consumer devices like phones, tablets, TVs, and hands- free speakers. In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound. Deep learning models can achieve state-of-the-art accuracy, sometimes exceeding human-level performance. Models are trained by using a large set of labeled data and neural network architectures that contain many layers.

Why Deep Learning matters

By using Deep learning methods we can achieve higher level of recognition accuracy. So, consumer electronics can meet user expectations, and it is crucial task for safety-critical applications like driverless cars. Recent advances in deep learning have improved to the point where deep learning outper- forms humans in some tasks like classifying objects in images. There are two main reasons it has only recently become useful:

□ Deep learning methods requires large amounts of labeled data, huge computational power.
Some Applications of deep learning
1) used in industries from automated driving to medical devices.
Automated Driving: In this researchers use deep learning to detect objects automatically and also to detect pedestrians to reduce accidents.
2) Aerospace and Defense: In this, by using Deep learning researches identify objects from satellites that are used to locate areas of interest, and identify safe or unsafe zones for troops.
3) Medical Research: Researchers can detect the cancer cell automatically by using Deep learning methods. An advanced microscope, made by UCLA Team takes a high- dimensional data set used to train a deep learning application to identify the cancer cells accurately.
4)Industrial Automation: In this, Deep learning is helpful to improve the safety of the worker around heavy machinery automatically when people or objects are within an unsafe distance of machines.
5) Electronics: Hearing and speech translation can be done automatically by using Deep learning methods. For example, in some home assistance devices, they respond to your voice and know your preferences are powered by deep learning applications.

 Which neural network can be opt for the training of video frames?

As explained above, about the artificial neural network and its types the neural network that can be well suited for training of labeled video frames is CONVOLUTIONAL NEURAL NETWORK

How Deep Learning works?

Deep learning models are trained by using large sets of labeled data and neural network architectures which are having more number of hidden layer to learn features directly from the data without the need for manual feature extraction.

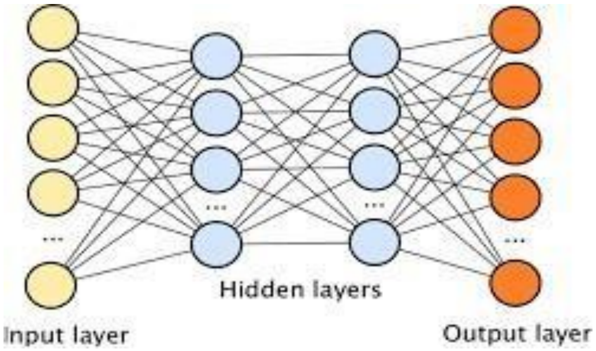

Figure 8: Deep learning network

Conclusion and Future work

Conclusion

All the data from both the languages kannada and telugu are gathered i.e; all the genres and the ten songs for each respective genre and are labelled as Vocals and Non vocals and verified for reducing the missing are mislabelled data and are corrected. As this work is purely based on Music which is perception oriented one we cannot conclude that it is perfectly classified or labelled based upon the work of each individual. The final part of the task is that all the data that is required for training the machine are collected and are manually classified and it is ready for the future work. The final labelled audio file will look like the below mentioned figure which has the labels.

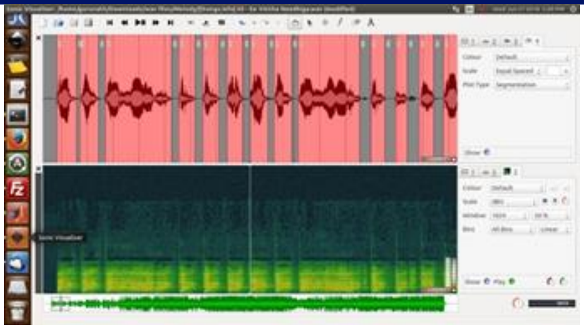

Figure 9: In the above figure 9 the labels are differentiated as grey and red colours by the sonic visualizer tool in order to have a clear difference.

Future Works

• The collected and organised data is taken as the training set and the validation set for training the machine.

• The data is given to the Neural network algorithm which will gettrain and learn the features from the given inputs and outputs.

• Here in this task of training the machine with our data we are not specifying are taking any features that should be learned by the data.

• But the algorithm or the process that we will adopt for this neural network training itself extract the features that it finds to be extracted from a file and learns them in order to classify the further files which will form our final out put.

• So the machine will uses the patterns or the features that which it find to be extracted instead of the manual specifying the features.

• This actually reduces the manual involvement and also the time will be saved as we are not manually extracting or specifying the features.

References

[1]Shayamal Patel,” Tools For Machine Learning For Use with MATLAB.” [online]. Available: https://in.mathworks.com/videos/essential-tools-for-machine- learning-1481139920800.html

[2]A. K. Jain, Jianchang Mao and K. M. Mohiuddin, ”Artificial neural networks: 1996. doi: a tutorial,” in Computer, vol. 29,

10.1109/2.485891 no. 3, pp. 31-44, Mar

[3]Shashank Easy For Prasanna, Use with "Machine MATLAB," Learning [online]. Made Available: https://www.mathworks.com/matlabcentral/profile/authors/2963954- shashank-prasanna.html

[4]J. Schmidhuber, "Deep learning in Neural Networks: An Overview," in Neural Networks, Vol. 61, pp. 85-117, 2015.

[5]A. Krizhevsky, et al., "imagenet Classification with Deep Convolu- tional Neural Networks," in Advances in neural information processing systems, pp. 1097-1105, 2012.

[6]J. Smith, "Machine Learning Using MATLAB,", Create Space Inde- pendent Publishing Platform, 2017.