

**DRAFT**

**MUSIC APPLICATION**

BY-

Anant Agarwal ( 16103033 , B3)

Prakhar Saxena (16103003, B3)

Rishabh Dhawan (16103213, B4)

Sonu Kumar Gupta (16103271, B4)

**Abstract**

Music assumes a significant job in human's day by day life and in the cutting edge trend setting innovations. As a rule, the client needs to confront the assignment of physically perusing through the playlist of melodies to choose. This venture centers around making an application to recommend tunes for client dependent on their state of mind by catching outward appearances. Outward appearance is a type of nonverbal correspondence. PC vision is an interdisciplinary field that helps passes on an abnormal state comprehension of computerized pictures or recordings to PCs. In this framework, PC vision segments are utilized to decide the client's feeling through outward appearances. When the feeling is perceived, the framework proposes a tune for that feeling, sparing a great deal of time for a client over choosing and playing tunes physically. Feeling Based Music Player likewise monitors stuffs like number of preferences for every tune, sorts melodies dependent on preferences every tune jumps on YouTube , and redesigns the tune that will be played unfailingly.

**INTRODUCTION**

Music listening is a standout amongst the most perplexing of human practices. Most basic practices have a conspicuous utility that can be conceivably followed to the useful thought processes of survival and multiplication. Besides, in the variety of apparently odd practices, couple of practices coordinate music for securing so much time, vitality, and cash. Music listening is a standout amongst the most well known relaxation exercises. Music is a pervasive friend to individuals' regular daily existences.

Why do people listen to music? In the course of recent decades, researchers have proposed various capacities that tuning in to music may satisfy. In any case, extraordinary hypothetical methodologies, various strategies, and various examples have left a heterogeneous picture with respect to the number and nature of melodic capacities. Additionally, there remains no understanding about the hidden components of these capacities. Main part examination recommended three unmistakable hidden measurements: People tune in to music to direct excitement and state of mind, to accomplish mindfulness, and as an outflow of social relatedness. The first and second measurements were made a decision to be considerably more significant than the third—an outcome that appears differently in relation to the possibility that music has advanced basically as a methods for social union and correspondence. The ramifications of these outcomes are talked about in light of speculations on the starting point and the usefulness of music tuning in and furthermore for the utilization of melodic boosts in every aspect of brain research and for research in music discernment.

**RELATED WORK**

1.EMOTION RECOGNITION

In emotional recognition of face a notable advancement has been observed in the field of neuroscience, cognitive and computational intelligence. It is also proved by Kharat and Dudul that about 55% effect of overall emotion expression is as facial expression which is contributed during social interactions.

Actually, facial muscle generates monetary adaptation in facial appearance which can be recapitulated by incorporating Action Units. The six common emotions have been considered as globally recognizable as the movements of muscle are very similar of these emotional expressions from the people from various region and society. Therefore, we have mainly concentrated on the automatic recognition of these six fundamental emotions.

In general, emotion recognition is a two steps procedure which involves extraction of significant features and classification. Feature extraction determines a set of independent attributes, which together can portray an expression of facial emotion. For classification in emotion recognition the features are mapped into either of various emotion classes like anger, happy, sad, disgust, surprise, etc . For the effectiveness calculation of a facial expression identification model both the group of feature attributes which have been taken for feature extraction and the classifier that is responsible classification are equivalently significant. For a badly picked collection of feature attributes, in some cases, even a smart classification mechanism is not able to produce an ideal outcome. Thus, for getting the high classification accuracy and qualitative outcome, picking of superior features will play a major role.

2. SONG CLASSIFICATION USING LYRICS

Automatic classification of music is an important and well researched task in music information retrieval (MIR) ,McKinney and Breebaart, 2003. Previous MIR work has primarily focused on classifying mood [Logan et al., 2004, Hu and Downie, 2010], genre [Mayer et al., 2008], annotations [Tingle et al., 2010, Nam et al., 2012], and artist [Knees et al., 2004]. Typically one or a combination of audio, lyrical, symbolic, and cultural data is used in machine learning algorithms for this task [McKay et al., 2010]. Genre classification by lyrics presents itself as a natural language processing (NLP) problem. In NLP the aim is to assign meaning and labels to text; here this equates to a genre classification of the lyrical text. Traditional approaches in text classification have utilised n-gram models and algorithms such as Support Vector Machines (SVMs), k-Nearest Neighbour (k-NN), and Naïve Bayes (NB). In recent years the use of deep learning methods such as recurrent neural networks (RNNs) or convolutional neural networks (CNNs) has produced superior results and represent an exciting breakthrough in NLP [Kalchbrenner et al., 2014, Kim, 2014]. Whilst linear and kernel models rely on good hand selected features, deep learning architectures attempt to prevent this by letting the model learn important features themselves. However, not much research has looked into the performance of these deep learning methods with respect to the genre classification task on lyrics.

**Methodology**

**1.Dataset**

In an attempt to improve the final model even more, the network will be trained on a larger set than the one described previously. Instead of 9000 pictures, training will be done with 20000 pictures from the FER-2013 dataset. Newly composed validation (2000 images) and test sets (1000 images) from the FER-2013 dataset (Kaggle)are used.

The Million Song Dataset is a freely-available collection of audio features and metadata for a million contemporary popular music tracks.

Its purposes are:

* To encourage research on algorithms that scale to commercial sizes
* To provide a reference dataset for evaluating research
* As a shortcut alternative to creating a large dataset with APIs (e.g. The Echo Nest's)
* To help new researchers get started in the MIR field

The core of the dataset is the feature analysis and metadata for one million songs, provided by The Echo Nest. The dataset does not include any audio, only the derived features.

The Million Song Dataset is also a cluster of complementary datasets contributed by the community:

SecondHandSongs dataset -> cover songs

musiXmatch dataset -> lyrics

Last.fm dataset -> song-level tags and similarity

Taste Profile subset -> user data

thisismyjam-to-MSD mapping -> more user data

tagtraum genre annotations -> genre labels

Top MAGD dataset -> more genre labels

**2. Emotion detection**

Article Detection using Haar feature based course classifiers is an effective thing recognizable proof strategy proposed by Paul Viola and Michael Jones in 2001. It is an AI based approach where a course work is set up from a lot of positive and negative pictures. It is then used to recognize inquiries in various pictures. Here we will work with face acknowledgment. At first, the figuring needs a huge amount of positive (pictures of playful faces) and negative (pictures of hopeless faces) to set up the classifier. By then we need to isolate features from it. For this, we apply each and every component on all the arrangement pictures. For every part, it finds as far as possible which will arrange the faces to positive and negative pictures. Regardless, plainly, there will be goofs or misclassifications. We select the features with least mix-up rate, which infers they are the features that best masterminds the playful face and desolate pictures.

A while later, for this they displayed Cascade of Classifiers. As opposed to applying all of the features on a window, cluster the features into different periods of classifiers and apply one-by-one. If a window crashes and burns the principle arrange, discard it. We don't consider leftover features on it. If it passes, apply the second period of features and continue with the strategy. The window which passes all stages is a face locale. It relies upon the Haar Wavelet procedure to explore pixels in the image into squares by limit. This uses AI frameworks to get an abnormal state of precision dependent on what is assigned "getting ready data". This uses "important picture" thoughts to process the "features" recognized. Haar Cascades use the Adaboost learning count which picks couple of noteworthy features from an immense set to give a profitable result of classifiers.

**3. Song classification**

The most important task in pattern classification is feature extraction and selection.

The different criteria for extracting healthy features are: salient, invariant and discriminatory. For modelling the model using machine learning for training we need to represent a document as a feature vector. The most used model is **Natural Language Processing** also known as bag-of-words model. In this model , initially all the different words that occur while training are collected and are stored with their occurrence. Vocabulary her is a set of unique words where order doesn’t matter.

Let S1 and S2 be 2 documents in the dataset:

S1 = “ India is a diverse country.”

S2 = “ I live in India.“

Based on these documents, the vocabulary could be written as :

Vocabulary={India : 2 , is : 1 , a : 1, diverse : 1 , country : 1, I*:* 1 , live : 1 , in : 1 }

**Table 1 :** Representation of words in 2 sample documents

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | india | is | a | diverse | country | i | live | in |
| XS1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| XS2 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Sum | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

Now a question may arise that what does 0s and 1s represent. Like are they absolute count or binary count(i.e. 1 if the word occur in the document else 0). It all depends on the probabilistic models that can be found further.

**Tokenization**

It is a process of breaking down a statement into individual words , punctuations are removed and all letters are converted to lowercase which are then used as an input for algorithms like NLP.

Eg. NASA launches its first satellite.

After tokenization – nasa , launches , its , first , satellite .

**Stop words**

These are are words which are most common words in a language. A statements without them is rear. Hence we must remove them from the vocabulary. If not removed they decrease the efficiency. Examples of stop words are: the , is , am , are , and , or , its , etc.

Eg: India is a diverse country.

After removal of stop words – ( india , diverse , country ) these are the words that will add to the vocabulary.

**Stemming and Lemmatization**

It is the process of converting a word into its root form. The first ever stemming model was developed by Martin F. Porter , hence this is known as **Porter stemmer.** Stemming can generate words with no maning (like : thus -> thu), so to obtain a grammatically right word , hence we use Lemmatization. It is a costly process and has a small impact on the results.

Eg : Robert likes to play football .

After stemming – robert , like , to , play , football

**N – grams**

In this model, instead of counting a single word as a token , we consider a sequence of words as a token. All the examples above were a simplest case of n-gram also called *unigram. In that* each token had exactly 1 word, 1 letter, or a symbol. Researchers found that n-gram with 4<=n<=8 yielded the highest accuracy.

Eg: Robert likes to play football.

Unigram – robert , likes , to , play , football

Bigram – robert likes , to play , football

Trigram – robert like to , play football

**3.Web Crawling**

BeautifulSoup is a Python library. It is utilized for parsing XML and HTML. It functions admirably in a joint effort with standard python libraries like urllib.

The benefits of BeautifulSoup is that it can parse HTML like a basic XML and return the qualities expected (content) to us with much ado.

A fast script to Crawl youtube. It pulls the outcomes from the principal page of YouTube results utilizing urllib, and prints every one of the connections of the recordings by parsing the page using BeautifulSoup.

def webCrawl:

    query = urllib.parse.quote(textToSearch)

    url = "<https://www.youtube.com/results?search_query=>" + query

    response = urllib.request.urlopen(url)

    html = response.read()

    soup = BeautifulSoup(html, 'html.parser')

    for vid in soup.findAll(attrs={'class':'yt-uix-tile-link'}):

        links.append('[https://www.youtube.com](https://www.youtube.com/)' + vid['href'])

Relating to the project the lyrical words of each song  (based upon the predicted emoji) from the classified folder (refer flow diagram) is passed to the function as a query to be searched. The results page is then crawled to find the links of the respected songs. To play the best-suited song, it is selected considering the most liked video song. During Crawl the video title and corresponding no of likes are stored in a csv file  and the one with the largest number of likes is the played .

def getStats(link):

        page = requests.get(link)

        likes = re.search("with (\d\*.\d\*.\d\*)", page.text).group(1)

        title = re.search("property=\"og:title\" content=\"([^\n]\*)", page.text).group(1)

        return (likes, title)

def writeinCsv:

   with open('abcde.csv', 'a') as csvFile:

        writer = csv.writer(csvFile)

        writer.writerow(row)

    csvFile.close()

**REFERENCES**

1. <http://www.cvauni.edu.vn/imgupload_dinhkem/file/Chuyen%20De%20HCI/3_%20Analysis%20of%20Emotion%20Recognition%20using%20facial%20expression-speech.pdf>
2. https://arxiv.org/ftp/arxiv/papers/1706/1706.01509.pdf
3. https://ieeexplore.ieee.org/document/4624313
4. <https://www.irjet.net/archives/V3/i2/IRJET-V3I284.pdf>
5. <https://www.semanticscholar.org/paper/Comparing-SVM-and-na%C3%AFve-Bayes-classifiers-for-text-Hassan-Rafi/501cfa39b2d94443ddbf4bfb7bb82c9845fa789c>
6. <https://www.semanticscholar.org/paper/Genre-classification-for-million-song-dataset-using-Hu-Ogihara/d12c2f65d3792cd8aa3b9a80046f52fd77744403>
7. <https://www.semanticscholar.org/paper/A-Bayesian-Approach-to-Filtering-Junk-E-Mail-Sahami-Dumais/4b7d5d9203e4ec2f03c1ffeff97fe28ead9d9db1>
8. [https://www.semanticscholar.org/paper/Automatic-Mood-Classification-Using-TF\*IDF-Based-on-Zaanen-Kanters/2086ead3dc0f4c9d9e8b193aa2871c8f8c782fa3](https://www.semanticscholar.org/paper/Automatic-Mood-Classification-Using-TF*IDF-Based-on-Zaanen-Kanters/2086ead3dc0f4c9d9e8b193aa2871c8f8c782fa3)
9. https://www.semanticscholar.org/paper/Comparing-SVM-and-na%C3%AFve-Bayes-classifiers-for-text-Hassan-Rafi/501cfa39b2d94443ddbf4bfb7bb82c9845fa789c