Emousic: Emotion and Activity-Based Music Player Using Machine Learning

Chapter · May 2019

DOI: 10.1007/978-981-13-6861-5_16

4 authors, including:

Jagannath Aghav
College of Engineering, Pune
34 PUBLICATIONS **100** CITATIONS

SEE PROFILE

Emousic : Emotion & Activity based Music Player using Machine Learning

Pranav Sarda[1*] ,Sushmita Halasawade[2], Anuja Padmawar[3] and Jagannath Aghav[4]

College of Engineering Pune, Wellsley Road, Pune, Maharashtra, India
[1] sardapranav1@gmail.com
[2] sushmita.halasawade@gmail.com
[3] anujapadmawar.ap@gmail.com
[4] jagannath.aghav@gmail.com

Abstract. In this paper we propose a new way of personalized music playlist generation. The mood is statistically inferred from various data sources primarily: audio, image, text, and sensors. Humans mood is identified from facial expression and speech tones. Physical activities can be detected by sensors that humans usually carry in form of cellphones. The state of the art data science techniques now make it computationally feasible to identify the actions based on very large data sets. The program learns from the data. Machine learning helps in classifying and predicting results using trained information. Using such techniques, applications can recognize or predict mood, activities for benefit to user.Emousic is s real time mood and activity recognition use case. It is a smart music player that keeps learning your listening habits and plays the song preferred by your past habits and mood, activities etc. It is a personalized playlist generator.

Keywords: Data Science, Emotion/Mood & activity recognition, Music analysis, Music playlist generator, classification.

1 Introduction

Smart gadgets people carry everyday with them can obtain lot of data. Data from fitness band, emotion detection from face captured by smartphone, activities detected by the sensors, further can be used for various applications. One of the very precise and straight approach to detect mood is using human facial expressions. Most of the time, emotion is revealed by face itself. Industries have their trained model for the emotion extraction and are now delivering customer services as frameworks providing API with help of cloud based services. We used Microsoft's cognitive servies & Google's own activity recognition API in our approach for faster implementation. Machine learning can be used for classifying music into set of particular emotions. Once, all this data is present, user can be studied about his/her preferences and habits of listening by time, mood, activity, etc. Training on this data can generate better playlist for future listening. Objective behind this work is to let daily factors get considered for better music recommendation.

2 Literature Survey

2.1 On emotion recognition

The mood is statistically inferred from various data sources primarily: audio, image, text and sensors. In paper [1] author used boosted tree classifier for emotion extraction from short video sequence using audio and video, classify them into 7 emotion categories. For audio feature extraction openSMILE[2] toolkit is used.This model gives better accuracy for three emotions viz. angry, happy and neutral. In[3] Author et. al. proposed an approach for analyzing the extracted facial features, with artificial neural network (ANN) used to classify those into six emotions viz. anger, happy, sad, disgust, surprise & fear. Gabor Wavelets & Markov random fields is used in [4] & [5] respectively.

Paper[6] is about emotion detection in voice from voice mail messages. Three differ ent types of training sets viz. PhoneShell messages, CallHome Corpus, Oasis database were used. Feature vectors are trained using HMM (hidden Markov models) emotion wise. They have also used Gaussian Mixture models and Zwickers model for loudness. Text independent method is presented in [7] for emotion recognition from speech. Hidden Markov model is proposed for the classification of speech emotions into six categories as disgust, fear, anger, joy, sad. In paper[8], author proposed GMVAR model as the statistical classifier for modeling the temporal structure of the data which is useful for speech emotion recognition. In [9], the author designed oneclassinoneneural (OCON) network[10] (i.e. for each of eight emotions different sub neural networks) for emotion recognition.Paper [11] presented emotion extraction on textual data. Author proposed the emotion detector algorithm which gives weight to each emotion word with the help of traversing and parsing the emotional ontology. Emotional class with the highest weightage is the final emotion of the corresponding text data. In [12], separate mixture model (SMM) is used to find the similarity between input sentence and EARs. Maximum probability shows the emotional state of the input sentence as happy, unhappy or neutral. Unsupervised emotion detection using semantic and syntactic relations is explored in [13].

In [14], author discussed a way to recognize the emotion of the user by using builtin sensors of mobile phone. They created a soft keyboard which uses data provided by these sensors to find the user's emotional state. The soft Keyboard uses the Multiresponse linear regression.Another approach of using mobile data statistics and activity is explained in [15].

2.2 Emotion in music and emotion based music player

Music plays vital role in our life, different kinds of music or songs have different impact on our lives. People like to listen music according to their mood, but how do we get if particular song belongs to particular category of emotion? Just like detecting human emotion, there are systems developed for music emotion recognition. There are various approaches for this, one of them discussed in [16]. Generally the songs are classified according to their metadata like title, singer etc. but to classify music into emotion set, music signal analysis is considered. Problem is approached using multi label classifi-

cation, also it is possible for one music to belong to various classes. SVM is the classifier used, for feature extraction MARSYAS is used with feature vector having 30 dimension. Simple confusion matrix is used to check the accuracy, precision and recall. Same topic is explored in [17].

Paper [18], is about mood based music player application's design. Overall system works with two main modules, one that extracts emotion and other one is music Audio feature extraction module. Output from these two modules is considered further in EmotionAudio recognition module. For emotion extraction, facial image is processed. Image is given as input after converting it to binary format. Viola \& Jones object de tection framework is used for detecting parts of face. According to them, facial points of eyes and mouth depict the emotions accurately. Support Vector Machine (SVM) is used as classifier. In Audio feature extraction module musicaudio signal is categorized into 8 types of mood pair for eg. sadanger using auditory toolbox. Once emotion is obtained, songs that maps with it from classified dataset in second module are taken. Randomizer generates the play list.

In [19], there is another similar approach discussed about mood extraction & mapping it with assumed mood wise classified dataset of songs. Here facial detection method is used again as their crucial factor but instead of shapes of eyes and lips, points on face are taken. They used SVM (Support Vector Machines) as their classification method. These surveys helped us how researches have achieved their own way of emotion recognition in human and music, determining activity recogniton as well, this motivated us to integrate such small factors to create a new way of generating playlist in music apps which in turn ultimately considers ouputs of such small emotion recognition modules and calculates result based on it. These small modules have their own working machine learning based model.

3 Problems in current systems

Though there exists few similar music players, they utilized machine learning in facial emotion recognition or emotion wise classification of songs. These two aims can be achieved with different number of ways. Our goal is not to build the same thing again, rather use outputs of different components for better playlist generation. Current solutions consists of typically capturing face and mapping the mood recognized with songs and generates the playlist. Some people like to listen songs preferred to their activites like while reading, walking, running, while using social media etc. Hence considering factors like user's activities, preference, timing, etc. Emousic is the solution.

4 Proposed Solution

4.1 Random Forest Classifier

Random forest is the supervised classification algorithm. It can be used in both classification and regression problems. Random forest builds many classification trees by

selecting best feature among a random subset of features. This process brings randomness, which results in a better model. In the classification process, each tree in the forest gives votes for that class. The forest chooses the classification having most of the votes. The general technique of bootstrap aggregating or bagging is applied in training algorithm for random forests.There is direct relation between the number of trees in the forest and the result we get, as larger the number of trees, better the accuracy. The advantages of Random forest algorithm is that it can handle the missing values, avoid the problem of overfitting and it can be modeled for categorical values.

4.2 Algorithm of Emousic's classifier

```
begin
  1. Fetch the dataset.
  2. Remove duplicate entries if present.
  3. If same set of independent variable has multiple dependent
  class then categorise them in one class.
  4. Divide the dataset into test set and training set
  5. Apply Random forest classifier to train the dataset.
  6. Take the inputs from data frame
  7. Predict the result as song by giving input to 5.
end
```

4.3 Architecture

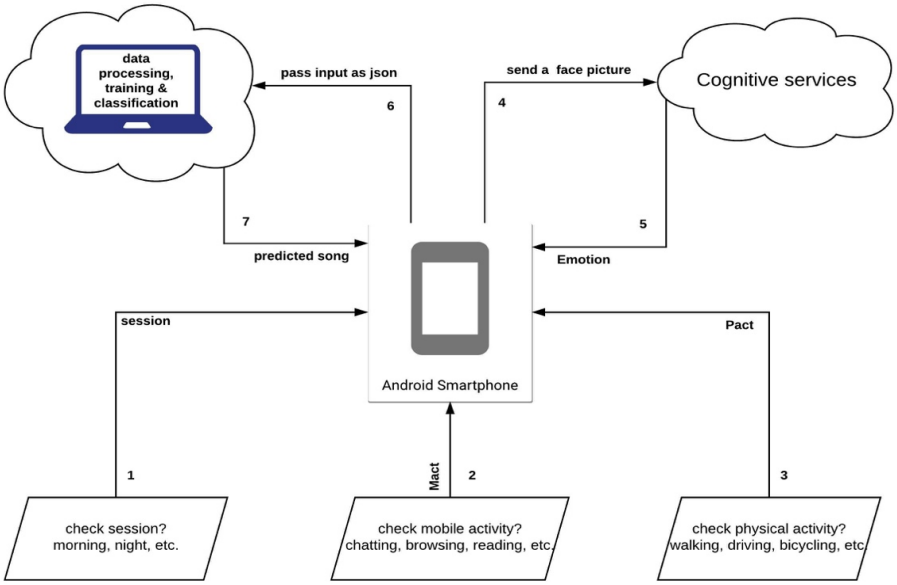

Fig. 1. Architecture

As shown in figure (1), System involves 3 modules that work independently altogether. Output from this modules acts as input to server which has trained model, model predicts the song and return back to phone. Modules are explained below.

Getting session -

This is simple function that detects if it is morning, afternoon, evening or night based on time and stores it into 'session' variable.

Module - 1 : Mobile Activity Detection

This module looks into application usage statistics of android and gets the current fore ground app being used by user other than music player itself.. Since we have considered few predefined categories as Social media, browsing, chatting, gaming, reading and if none of these, then nothing. Say for e.g., user just opened player after/while using WhatsApp, it will be stored as "chatting" in the 'Mact' variable.

Module - 2 : Physical Activity Detection

This module detects physical activity from the data collected from smartphone sensors. Google has already trained their model and provide service in form of API. This activities are classified into categories like Driving, Walking, Running, Still, Bicycling. It stores them into 'Pact' variable.

Module -3 : Emotion Recognition System

This module takes input as user's selfie and pass it to cognitive services API provides by Microsoft for emotional analysis. We used this API to build our project in short time, however one can implement his own model for facial emotion recognition and use it as a module. This API, classifies emotions into eight emotion categories viz. Happy, Sad, Angry, Surprise, Neutral, Contempt, Disgust, Fear.

Output of these 3 modules and session are stored into JSON object and passed on to Server. Server runs a python script which accepts this JSON objects fields into variables and pass it to R script as command line arguments to run classifier script. R script then computes and predicts the best suitable song and prints out song name in form of JSON. This STDOUT is collected by python script and sent back to android smartphone. Cursor points to the directory of emotion wise classified songs and plays the song returned as JSON. Till this, it is completely automated, then user can take full control of application.

5 Dataset

Since this approach is used for personalization which varies from user to user, dataset varies too. We used one smartphone and dumped output of emotion, physical activity, mobile activity and session and songs listened by phone owner periodically in the phone itself. Later this database in csv format was used in server side for training. There are 4 independent variables i.e. mood, physical activity(pact),

mobile activity(mact), session and one dependent variabe i.e. song. Each independent variable is categorically defined. Mood contains happy, sad, fear, anger, contempt, disgust, neutral, surprise (output by Microsoft's cognitive services). Pact has values as driving, walking, running, still, bicycling (output by Google's activity recognition API). Mact contains social media, chatting, gaming, reading, nothing. Session is daytime as morning, afternoon, evening and night. Experimentation was performed on 101 songs. These songs are manually tagged into their emotion sets.

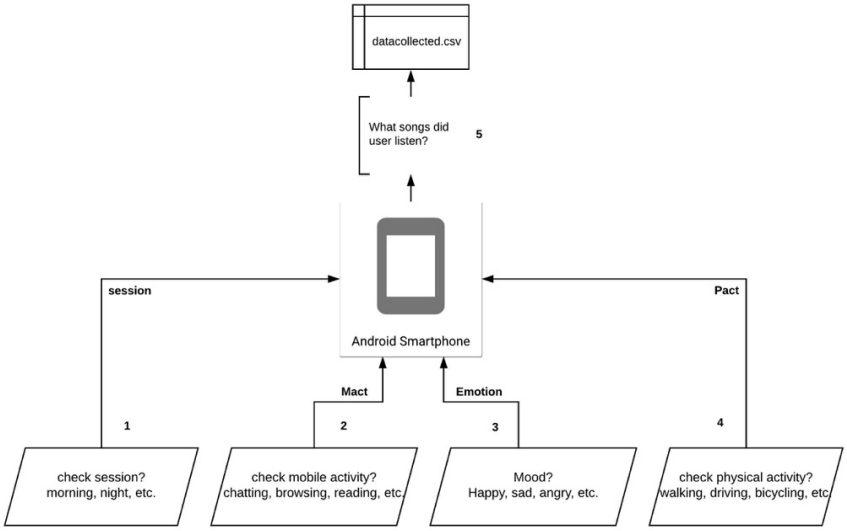

Fig. 3. Data Collection Model

6 Experimentation

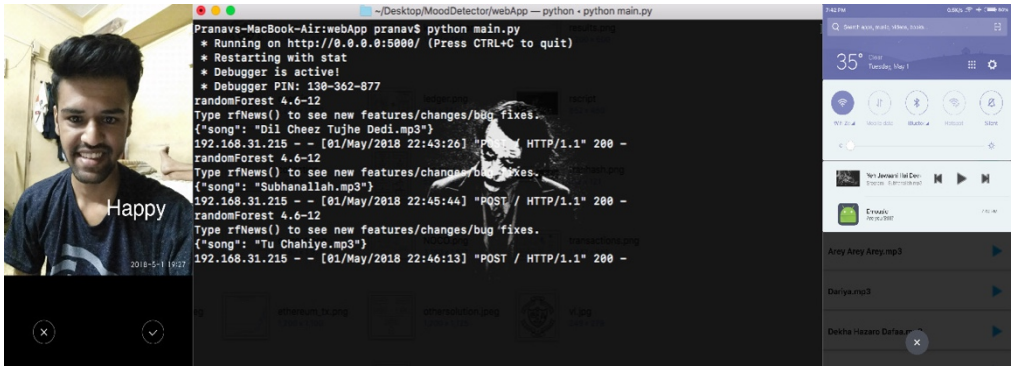

Fig. 3. Emousic App demo

Above explained architecture was implemented in the android application. Microsoft's cognitive services for mood and Google's activity recognition API for physical activity recogniton were used for faster implementation. It works as explained in figure (4). All modules run parallely, output of each module is stored in variables and this as input is passed in form of JSON object to webservice for getting recommended song. Web-service then receives this input, it runs classifier prediction and build a JSON containing song name and send it back to phone. Emousic then autoplay this song and show playlist of all songs mapped to its emotion.

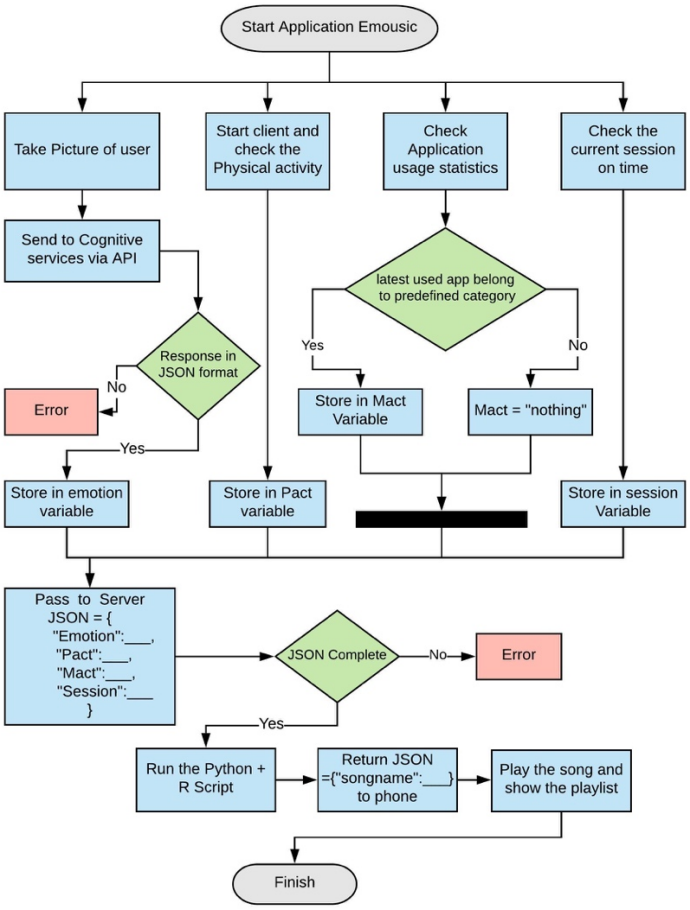

Fig. 4. Flowchart of Emousic

7 Results

Parameters	J48 Algorithm	Random Forest
Correctly Classified instances	50.86%	96.33%
Incorrectly Classified instances	49.14%	3.67%

We trained the dataset which had approximately 1000 entries in training set and 100 entries in the test set. Weka tool is used to find the accuracy of different machine learning algorithms used to train the model. From the above table it is observed that, Random forest algorithm gives better accuracy than J48 algorithm with the mean absolute error of 0.0066 and 0.014 respectively. The accuracy of random forest algorithm can be incresed by increasing the number of trees in the forest.

8 Conclusion and Future Scope

This paper presents Emousic, a new way of personalizing songs playlist by using machine learning techniques. Our solution works and gives better user preferable playlist. Due to the large number of classes, the performace of Random forest classifier is better than Decision tree algorithms. Since the experiment was performed on a small dataset and limited number of features, it still can be improved by adding more features like age, weather etc. More number of attributes will improvise decision making and prediction of song. Each user has it's own preferences about what kind of song is to be played for corresponding mood. for e.g., some users listen sad songs when they are sad while some may prefer happy songs to change their mood.

Collecting this data from every user can help us build better user specific radio application. Implementing this prototype in current music applications can provide better music experince to user.

References

1. Day, Matthew. Emotion recognition with boosted tree classifiers. ICMI 2013 Proceedings of the 2013 ACM International Conference on Multimodal Interaction. 531-534. 10.1145/2522848.2531740.
2. Eyben F., Wllmer M. and Schuller B. openSMILE the Munich versatile and fast opensource audio feature extractor. In Proc. ACM Multimedia. 1459-1462.
3. Renuka R. Londhe, Vrushsen P. Pawar. Analysis of Facial Expression using LBP and Artificial Neural Network International Journal of Computer Applications (0975-8887),Volume 44-No.21, April 2012.
4. Michael lyon, Shigeru Akamatsu. Coding Facial expression with Gabor wavelets. IEEE conf. on Automatic face and gesture recognition, March 2000.
5. Maglogiannis, Ilias and Vouyioukas, Demosthenes and Aggelopoulos, Chris. Face detection and recognition of natural human emotion using Markov random fields. Personal and Ubiquitous Computing,volume 13, pages 95-101, Jan 2009.

6. Zeynep Inanoglu, Ron Caneel. Emotive Alert: HMM-Based Emotion Detection In Voicemail Messages. MIT Media Lab Technical Report No. 585, January 2005. Appeared in:Intelligent user Interfaces (IUI 05), 2005, San Diego, California, USA.
7. Tin Lay Nwe, Say Wei Foo, Liyanage C.De Silva. Speech emotion recognition using hidden Markov models. Elsevier Speech Communications Journal Vol. 41, Issue 4, pp.603-623, November 2003.
8. Moataz M. H. El Ayadi, Mohamed S. Kamel, Fakhri Karray. Speech Emotion Recognition using Gaussian Mixture Vector Autoregressive Models. IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007.
9. J. Nicholson, K. Takahashi, R. Nakatsu. Emotion Recognition in Speech Using Neural Networks. Neural Computing and Applications, volume 9, pages 290-296, December 2000, ISSN 1433-3058.
10. Markel JM, Gray AH. Linear Prediction of Speech. SpringerVerlag, 1976.
11. Shiv Naresh Shivhare, Saritha Khethawat. Emotion Detection from Text. CoRR, volume abs/1205.4944, 2012.
12. Wu, ChungHsien, ZeJing Chuang, YuChung Lin. Emotion recognition from text using semantic labels and separable mixture models. ACM Trans. Asian Lang. Inf. Process., volume 5, pages 165-183, 2006.
13. Ameeta Agrawal, Aijun An. Unsupervised Emotion Detection from Text using Semantic and Syntactic Relations. 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, Macau, 2012, pp.346-353.
14. Zualkernan, F. Aloul, S. Shapsough, A. Hesham, Y. ElKhorzaty. Emotion recognition using mobile phones. Computers and Electrical Engineering 60 (2017), page 113.
15. Kiran K. Rachuri, Mirco Musolesi, Cecilia Mascolo, Peter J. Rentfrow, Chris Longworth, Andrius Aucinas.(2010). EmotionSense: A Mobile Phones based Adaptive Plat-form for Experimental Social Psychology Research. 281-290. 10.1145/1864349.1864393.
16. Tao Li, Mitsunori Ogihara. Detecting Emotion in Music. ISMIR (International Conference on Music Information Retrieval),2003.
17. Namrata Mahajan, Harshad Mahajan. Detecting Emotion in Music. International Journal of Electrical and Electronics Research ISSN 2348- 6988 (Vol. 2), Issue 2, pp: (56-60), Month: April - June 2014.
18. Hafeez Kabani, Sharik Khan, Omar Khan, Shabana Tadvi. Emotion Based MusicPlayer. International Journal of Engineering Research and General Science Volume 3, Issue 1, January-February, 2015.
19. Abhishek R. Patel, Anusha Vollal, Pradnyesh B. Kadam, Shikha Yadav, Rahul M. Samant. MoodyPlayer: A Mood based Music Player. International Journal of Computer Applications (0975 8887) Volume 141 No.4, 2016.