ISSN 2395-1621

Music Genre Classification using Hybrid Deep Learning Approach

Pratik Bhagwat, Chetan Dobhada, Prathamesh Khandake, Rushikesh Walke

p.bhagwat@rocketmail.com,
chetan.dobhada0608@gmail.com,
prkhandake@gmail.com,
rushikeshwalke77@gmail.com

Department of Computer Engineering
Sinhgad Institute of Technology & Science, Pune-411041

ABSTRACT

Musical genres are categorical descriptions that are used to describe music. They are commonly used to structure the increasing amounts of music available in digital form on the Web and are important for music information retrieval. Genre categorization for audio has traditionally been performed manually. A particular musical genre is characterized by statistical properties related to the instrumentation, rhythmic structure and form of its members. In this work, CNN and Bi-RNN algorithms and it's fusion for the automatic genre categorization of audio signals are described. More specifically, proposed set of features for representing texture and instrumentation. In addition a novel set of features for representing rhythmic structure and strength is proposed. The performance of those feature sets has been evaluated by training statistical pattern recognition classifiers using real world audio collections. Based on the automatic hierarchical genre classification two graphical user interfaces for browsing and interacting with large audio collections have been developed.

ARTICLE INFO

Article History

Received: 26th September 2020

Received in revised form :

26th September 2020

Accepted:28th September 2020

Published online :

28th September **2020**

Keywords: Music Genre, Categorization, CNN, Bi-RNN

I. INTRODUCTION

Musical genres are categorical descriptions that are used to describe music. They are commonly used to structure the increasing amounts of music available in digital form on the Web and are important for music information retrieval. Genre categorization for audio has traditionally been performed manually. A particular musical genre is characterized by statistical properties related to the instrumentation, rhythmic structure and form of its members. In this work, algorithms for the automatic genre categorization of audio signals are described.

More specifically, we propose a set of features for representing texture and instrumentation. In addition a novel set of features for representing rhythmic structure and strength is proposed. The performance of those feature sets has been evaluated by training statistical pattern recognition classifiers using real world audio collections. Based on the automatic hierarchical genre classification two graphical user interfaces for browsing and interacting with large audio collections have been developed. Automatic music genre classification is a widely explored topic. However, almost all related work is concentrated in the classification of music items into broad genres (e.g., Pop, Rock) using handcrafted audio features and assigning a single label per item. Even though there has been some developments in this area, the accuracy is still an issue in the genre based classification and music recommendation system. Hence it will be easier to use modern deep learning techniques to implement the classification task and recommend better music based on genre.

Everyday more and more people are tuning in to listen to their favorite music. They are especially an audience to the web music where they get access to different types of music at once. Though music has different aspects such as mood based, theme based, genre based it is difficult to classify data this large in a single field. Hence, we use different ML techniques to make this task easier and provide a better service which will also be beneficial to the money makers and service providers.

Librosa [16] is a Python package for audio and music signal processing. Version 0.4.0 is used for the audio processing. This is because librosa provides assistance for implementing multiple common functions used for the retrieval of complex music information. Even though there have been some developments in this area, the accuracy is still an issue in the genre based classification and music recommendation system. Hence it will be easier to use modern deep learning techniques to implement the classification task and recommend better music based on genre.

II. LITERATURE SURVEY

This chapter tells about the description of this project. It gives an idea about how the project is distributed in parts and the techniques that will be used to implement the project. This chapter includes the related work studied in relation with this project. These papers are close to the objectives of the project and the observations of these research papers are analyzed in the project. The literature survey is divided into 3 main themes. In the first section, traditional methods were analyzed but these methods did not give perfect results. So Ensembled methods were introduced to give better output by combining different traditional methods. In the third section, Deep learning methods gave more enhanced results than previous strategies were studied.

1. Critical Survey on Traditional Machine Learning Approaches

Changsheng Xu,et.al.[5] has studied the `Musical genre classification using support vector machines'. The paper summarizes the automatic musical genre classification is very useful for music indexing and retrieval, because the author feels that the amount of digital music increases rapidly nowadays, to effectively organize and process such large variety and quantity of musical data to allow efficient indexing, searching and retrieval is a big challenge. The authors have studied the multi-layer classifier based on support vector machines (SVM) methodology. The authors claim that it has good performance in musical genre classification and are more advantageous than traditional and other methods. The authors got the results (error rate) using SVM which was 6.86 percent which was very minimum with compares to other methods. But the limitations in that Support vector machines take a long time in the training process, especially with a large number of training samples.

Carlos N. Silla Jr., et.al. [6] Used an ensemble approach according to time and space decompositions: feature vectors are selected from different time segments of the music, and one-against-all and round-robin composition schemes are employed for space decomposition which show that the employed features have different importance according to the part of the music signal from where the feature vectors were extracted. Furthermore, the ensemble approach provides better results than the individual segments in most cases. Also, the use of a reduced set of features implies a smaller processing time. This point is an important issue in practical applications, where an adequate compromise between the quality of a solution and the time to obtain it must be achieved

Takuya Kobayashi,et.al.[7] have proposed novel audio features using correlations between sub-band signals and showed its effectiveness for music genre classification. The proposed method demonstrated the best accuracy of 81.5 percent, outperforming the conventional methods. In future, we can use other training/classification method such as linear discriminate analysis for further Classification Accuracy (CA) improvement. We can also reduce the dimension of the features without degrading CA by selecting effective sub-band correlations from all the combinations. Gaussian Processes (GPs) having capabilities to identify nonlinear data relations such as time series analysis and classification tasks.

K. Markov and Tomoko Matsui [8] has tried to use the applicability of GP models for music genre classification and emotion estimation. Here, two systems are operating, one for music genre classification and another for music emotion estimation using both SVM and GP models. The music was processed in the same way and the effect of different feature extraction methods and their various combinations were also observed. It was observed that in both the tasks, GP performed consistently better than SVM. GP produce Gaussian distribution as their output in contrast to SVM which provides sparse solution.

Bob L. Sturm [12] describes the bibliography of work in MGR and analyzes three aspects of evaluation like experimental designs, datasets, and figures of merit. They present summary statistics of each. In the experimental designs they explained ten designs of MGR and their accuracy. They found that classify is the most accurate method for MGR. The author did survey of the datasets that were used for MGR. In this paper, figures of merit have been briefly described. Mariusz et.al. [15] examined the utilization of Sparse Autoencoders (SAE) in the process of music genre recognition. Scattering Wavelet Transform (SWT) has been used as an initial signal representation. The SWT uses a sequence of Wavelet Transforms to compute the modulation spectrum coefficient of multiple order switch was already shown to be promising for this task. The Auto encoder scan be used for pre-training a deep neural network, treated as an features detector, or used for dimensionality reduction. In this paper, SAE's were used for pre-training deep neural network on the data obtained from jamendo.com website offering music on creative commons licence. The pre-training phase is performed in unsupervised manner. Using a simple Sparse Autoencoder, the author improves the result even in the simplest case and even outperforms the best MLP by a little margin. The authors suspect that the normalization of the feature space plays a role in the adaptability of the SA.

Chih-Hsun Chou, et.al [14] have used spectrogram analysis to analyze the characteristics and genre classification of the music. In the proposed method, two important methods were integrated to extract the desired features. The capability of multi resolution analysis of the wavelet package decomposition was integrated with dimension reduction ability of the singular value decomposition. Experimental results with the well-known ISMIR 2004 and GTZAN database were used to verify the performance of the proposed method. Music was transformed into sixty two sub-bands by five levels of WPD so that the spectrogram of each sub-band could be obtained using short time Fourier transformation.

2. Critical Survey on Ensembled approaches

Paulo Ricardo, Lisboa de Almeida et.al [9] has presented a dynamic ensemble selection method for music genre classification, where two pools of diverse classifiers are created by using different features types. These feature types are extracted from different music pieces. By using the k-nearest oracles method, the classifiers are ensembled to dynamically select the test patterns. As an output the model efficiently gave about 63 to 70 per cent accuracy. Further, the weaker classifiers can be replaced by the pools composed of SVM to test other strategies to select the classifiers from the pool.

Gjorgji Madjarov, Goran Pesanski, et.al [10] have explored the task of automatic music genre classification. Multiple features based on timbral texture, rhythmic content and pitch content are extracted from a single music piece and used to train different classifiers for genre prediction. For the classification, two different architectures flat and hierarchical classification and three different classifiers (kNN, MLP and SVM) were tried. .The experiments carried out on a large dataset containing more than 1700 music samples from ten different music genres have shown accuracy of 69.1% for the flat classification architecture (utilizing one against all SVM based classifiers). The accuracy obtained using the hierarchical classification architecture was slightly lower 68.8%, but four times faster than the flat architecture. Future work will involve further analysis of the feature space, genre group dependent selective extraction and combination of different types of features on the second level of the classification hierarchy, examination of alternative classification schemes, and incorporation of more audio classes

Loris Nanni, et.al.[11] studied the Music Information Retrieval (MIR) system. In this work the author, present the novel and effective approach for automated musical genre recognition based on the fusion of different set of features. Both acoustic and visual features are considered, evaluated, compared and fused in a final ensemble, which show classification accuracy comparable or even better than other state of the art approaches. The music genre classification system combines audio and visual features. In this paper 11 different texture descriptors extracted from the spectrogram image and several acoustic feature vectors are evaluated and compared. The combined approach has a main drawback which is the increased computational cost needed for feature extraction. With respect to existing approaches based on audio features the proposed approach introduces a big innovation. It shows that an audio signal can be represented using a visual representation and that visual features have a great discriminant power in music genre classification. This assertion opens new research directions since the number of textural features proposed in the literature that can be tested for this classification problem are very large.

Mckay, et.al[17] find out that lot of expertise and time is required to manually classify recordings and also there is limited agreement among human annotators when classifying music by genre, so they used Gaussian Mixture Models (GMMs), obtaining a maximum of 99% recognition.A database was elaborated to train the classifiers, and an accuracy of 98% was achieved when classifying among four styles. When using eight classifiers, trained to return ‘‘yes’’ or ‘‘no’’ for eight different styles, they got an accuracy of 77–90%

3. Critical Survey on Deep Learning Approaches

Vishnupriya S, et.al [2, 3] had used Convolution neural network for training and classification of Music. CNN classifies music into various genres by extracting the feature vector. Results show that the accuracy level of system is around 76 percent and it will greatly improve and facilitate automatic classification of music genres. Also with the use of CNN along with small set of 8 music features having 3 main music dimensions results more efficiency of 89 percent in Music genre classification. The future work will focus on developing the system further to classify the songs based on mood. In addition, it can be possible to increase the accuracy and efficiency of classification by having a Fusion of two different neural network models.

Keunwoo Choi, et.al. [4] Had used CRNN architecture for music classification, which takes advantage of convolution neural networks (CNNs) for local feature extraction and recurrent neural networks for temporal summarization of the extracted features. Overall, they found that CRNNs show strong performance with respect to the number of parameter and training time, indicating the effectiveness of its hybrid structure in music feature extraction and feature summarization. Future work will investigate RNN-based structures and audio input requirements for deep learning approaches.

Sergio Oramas, et.al [13] have proposed an approach to learn and combine multimodal data representations for music genre classification. Intermediate representations of deep neural networks are learned from audio tracks, text reviews, cover art images, and further combined for classification. Experiments on single and multi-label genre classification are then carried out, evaluating the effect of the different learned representations and their combinations. Results on both experiments show how the aggregation of learned representations from different modalities improves the accuracy of the classification, suggesting that different modalities embed complementary information. In addition, the learning of a multimodal feature space increases the performance of pure audio representations, which implies a more fine-grained categorization. In addition, an approach is proposed based on the learning of a multimodal feature space and a dimensionality reduction of target labels using PPMI. Results show in both scenarios that the combination of learned data representations from different modalities yields better results than any of the modalities in isolation.

Hence, from the above literature survey it was observed that the traditional methods had some drawbacks that were improved by Ensembled approaches still it did not give the expected accuracy. When Deep learning approaches were used for Music genre Classification it has shown better accuracy but for expected results we could use hybrid deep learning approach for the classification

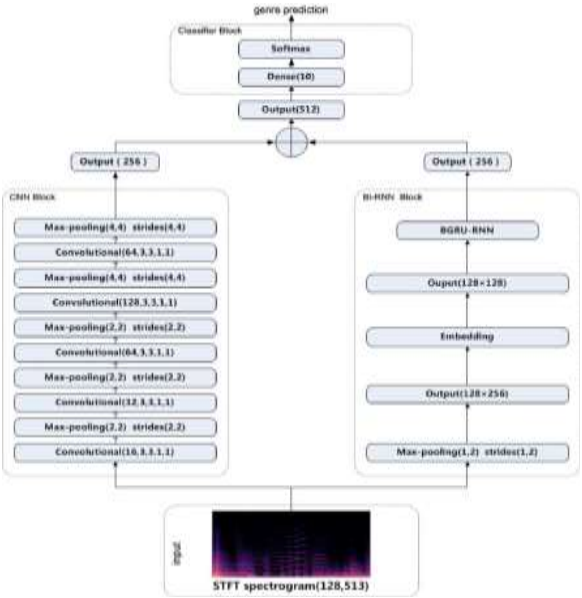
Methodology:-

As the current Music genre classification system doesn’t provide playlist suggestions based on genre. Hence in this system, we propose architecture to improve the performance of music genre classification & also suggest the playlist based on said genre. This end-to-end model consists of paralleling CNN and Bi-RNN for feature extraction.The outputs of two paralleling blocks are fused into a more powerful feature vector for music classification. In this complete system, we had used PyQt for UI designing, Python for developing business logic, MySQL for database purposes,Tenser flow for Neural networks implementation, Librosa for music information retrival, Pyaudio for music/audio processings.

In this methodology, we had used hybridized architecture of deep learning approaches like Convolutional Neural Networks (CNN) and Bidirectional Recurrent Neural Network. In which prediction is made on unknown music sample set using CNN and is optimized by using Bi-RNN. For that, Train, validation and test sets are created and samples are taken from dataset. Next, the convolutional layer of algorithm will Input music and its features are detected and are mapped. Next, Rectied Liner Unit function is applied to increase the non-linearity (removes noisy and

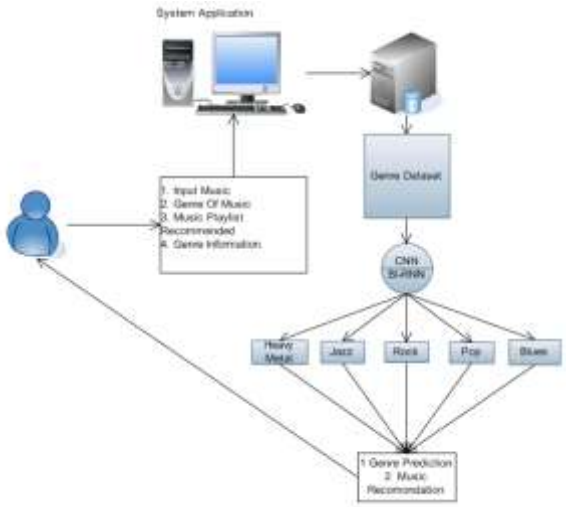
uncertain values in the input) in the feature map. after that, feature map is taken as input by next layer i.e. Max Pooling which reduces the number of parameters within the model and pooled map is generated Pooled Map is attened into a column like structure which is long vector of input data that you then pass through the articial neural network to have it processed further. It takes this data and combine the features into a wider variety of attributes that make the convolutional network more capable of classifying Music.After this, the network formed is compiled and trained using dataset.

Unknown music sample is evaluated on this convolutional neural network and for the optimization of the output of CNN, it is processed by a max-pooling layer of Bi-RNN to reduce the dimension. Further in Bi-RNN, the dimension of spectrogram is reduced to 128 x 256. Since the upper BGRU layers are constructed kinder complex, which employ an embedding layer for further dimension reduction to decrease parameters. After the pre-training, a 128 x 128 input is fed into two stacked BGRUs illustrated for features extraction. In this the, splicing of the outputs of two stacked BGRU layers as one 256D feature vector. Which is processed to give an optimized prediction for genre of the music sample.



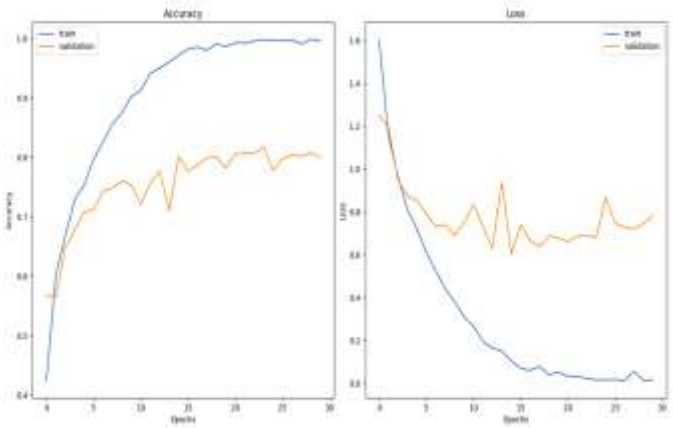
III. SYSTEM ARCHITECTURE

As shown below system design the overall design of system lies. It has music files as input and it predicts the genre of music as output. Also system will recommend the similar genre music playlist to user by assuming that the user's current mood is for listening the same genre. For all this purpose the system will have its own dataset of genres where already sorted genre and their music file will be stored. System then preprocess this dataset and will train itself for predictions using deep learning algorithms such as CNN and Bi-RNN.After training, system will test model generated after train and check ground truth and predicted values to check accuracy of the system as well. These roles has to be perform by admin only to keep system updated. User will give simply input sample of music whose genre is unknown for user. System will then map this sample with the model trained and predict a genre for the _le given input. The similar genre playlists will be stored in MySQL database and accordingly and user will be provided with a playlist of music files of the same genre.Also basic information of genre will be shared with user by retrieving from database.



IV. RESULTS

It was observed that the CNN model when integrated with Bi-RNN gives the best results with highest accuracy and lower loss. CNN based classification model was used to classify the genre set and extract the features. In this python project, we implemented a CNN & Bi-RNN hybrid model to classify the song into it's respective genres.



Loss and Accuracy Graph

V. CONCLUSION

The revolutionary development in the field of machine learning where deep learning approaches were introduced, gave an opportunity to make a breakthrough in the field of classi_cation. Genre based music classification and music recommendation system is an important issue where development was necessary which can be achieved by using different deep learning techniques. It was necessary that the functionality of the system must be user friendly and accurate. In this project, a better deep learning approach for music genre classi_cation was targeted. By using feature extraction method, CNN algorithm gave a better result when paralleled with Bi-RNN algorithm. It was observed that Bi-RNN when integrated with CNN gave an e_cient model with comparatively higher accuracy. The dataset was

successfully trained through CNN algorithm and the model was further optimized by integrating it with the Bi-RNN algorithm. The training accuracy was about 95 percentages and the CNN model gave expected accuracy which was above 85 percentages.

REFERENCES

[1] Lin Feng, Shenlan Liu, Jianing Yao, "Music Genre Classification with Paralleling Recurrent Convolution Neural Network", December 2017

[2] Vishnupriya S K.Meenakshi, "Automatic Music Genre Classification using Convolution Neural Networks", 2018 International Conference on Computer Communication and Informatics (ICCCI -2017), 2018

[3] Christine Senac, Thomas Pellegrini, Florian Mouret, Julien Pinquier, "Music Feature Maps with Convolution Neural Networks for Music Genre Classification"

[4] Keunwoo Choi, Gyorgy Fazekas, Mark Sandler, Kyunghyun Cho, "Convolution Recurrent Neural network for Music Classification", ICASSP-2017

[5] Changsheng Xu, Namunu C. Maddage, Xi Shao, Fang Cao, Qi Tian, "Music Genre Classification using Support Vector Machine", Conference Paper in Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on _ May 2003

[6] Carlos N. Silla, Alessandro L. Koerich, Celso A. A. Kaestner, "A Machine Learning Approach to Automatic Music Genre Classification", September 2008

[7] Takuya Kobayashi, Yusuke Suzuki, Akira Kubota, "Audio feature extraction based on sub-band signal correlations for music genre classification", 2018 IEEE International Symposium on Multimedia (ISM), 978-1-5386-6857-3/18/31.00 c 2018 IEEE DOI 10.1109/ISM.2018.00-15

[8] Konstantin Markov and Tomoko Matsui, "Music Genre and Emotion Recognition Using Gaussian Processes", IEEE Access 2:688-697 _ January 2014

[9] Paulo Ricardo, Lisboa de Almeida, Eunelson José da Silva Júnior, Luis Eduardo Soares de Oliveira, Tatiana Montes Celinski, Alessandro Lameiras Koerich, "Music Genre Classification using Dynamic Selection of Ensemble of Classifiers", IEEE International Conference on Systems, Man, and Cybernetics October 14-17, COEX, Seoul, Korea, 2012

[10] Gjorgji Madjarov, Goran Pesanski, Daniel Spasovski, and Dejan Gjorgjevikj, "Automatic Music Classification into Genres", ICT Innovations Web Proceedings ISSN 1857-7288, 2012.

[11] Loris Nannia, Yandre M.G.Costab, Alessandra Luminic, Moo Young Kimd, SeungRyul Baek, "Combining visual and acoustic features for music genre classification", 2015 Elsevier Ltd.

[12] Bob L. Sturm, "A Survey of Evaluation in Music Genre Recognition", Springer International Publishing Switzerland 2014

[13] Sergio Oramas, Francesco Barbieri, Oriol Nieto and Xavier Serra, "Multimodal Deep Learning for Music Genre Classification", Transactions of the International Society for Music Information Retrieval, 2018.

[14] Chih-Hsun Chou, Bo-Jun Liao, "Music Genre Classification by Analyzing the Subband Spectrogram", IEEE, 2014

[15] Mariusz Kleca, Danijel Korzineka, "Unsupervised Feature Pre-training of the Scattering Wavelet Transform for Musical Genre Recognition", International workshop on Innovations in Information and Communication Science and Technology, IICST 2014.

[16] Brian McFee, Colin Raffel, Dawen Liang, Daniel P.W. Ellis, Matt McVicar, Eric Battenbergk, Oriol Nieto, "librosa: Audio and Music Signal Analysis in Python", PROC. OF THE 14th PYTHON IN SCIENCE CONF. (SCIPY 2015), 2015

[17] McKay C, Fujinaga. "Musical genre classification: is it worth pursuing and how can it be improved?" In: 7th Int conf on music, information retrieval (ISMIR-06); 2006.

