Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 4, Issue 11, November -2017

# SURVEY on Music Genre Recognition using Deep Learning

Maulik Desai[1] , Jay Rodge[2], Kapil Sahu[3], Advait Kulkarni[4] , Prof. Bhavana Bahikar[5]

[1,2,3,4,5] *SKNSITS, Lonavala*

**Abstract:** *Music genre is clear cut names made by people to classes bits of music. A music genre classification is portrayed by the basic attributes shared by its individuals. These attributes ordinarily are identified with the instrumentation, harmonic content, and rhythmic structure of the music. Genre is usually used to structure the expansive accumulations of music accessible on the Web. Presently, music genre annotation is performed manually. Automatic music genre classification arrangement can help or supplant the human client in this procedure and would be a valuable expansion to music data retrieval frameworks. Likewise, Automatic music genre classification gives a structure to creating and assessing highlights for a substance based examination of musical signals. In this survey paper, the automatic music genre into a progressive system of music genre is investigated. All the more particularly, three feature sets for representing pitch content, rhythmic content and timbral texture are proposed. The performance and relative importance of the proposed features are investigated by training statistical pattern recognition classifiers using real-world audio collections. We compare the classification accuracy rate of various deep learning models with a set of well-known learning models including Deep Neural Network, Convolution Neural Network, Recurrent Neural Network in combination with hand-crafted audio features for a genre classification task on a public dataset.*

***Keywords***- *Deep Learning; Cloud Computing; Recurrent, Music Decomposition, Music Genre Recognition, Tag Retrieval*

## I INTRODUCTION

CNNs have been used extensively in solving various complicated machine learning problems such as sentiment analysis, feature extraction, genre classification and prediction. Hybrid models of CNNs and RNNs have been recently applied for temporal data like audio signals and word sequencing. Convolution Recurrent Neural Networks (CRNN's) are complex neural networks formed by combining Convolution CNN and RNN networks. CRNN architecture as a modified model of CNN with a RNN structure placed over it. This architecture has the capability to be as a robust structure to extract local feature using CNN layers and temporal summation by RNN networks. CNN's have been very popular in music recognition in diver's aspects such as automatic tagging, hybrid music recommender and feature learning. The key elements for a CNN network are: type of input signal, learning rate, activation function, batches and architecture. Mel-spectrogram is the preferred input type for music information retrieval. Mel-spectrograms consist of widespread fe1atures for tagging, boundary and onset detection, latent feature learning and it has been proved that Mel-scale is similar to the human auditory system. To achieve mel-spectogam signal, STFT (short time Fourier transform), and Log-amplitude spectrogram are required as pre-processing phase. Music feature learning with deep networks was improved with ReLu as activation function. Later this function is replaced with ELU (Exponential Linear Unit) to get fast and accurate learning. Recurrent neural networks also experienced significant improvement when gated recurrent neural network are applied. Gated RNN's have gating units which limit the flow of information through them, allowing to capture critical information from different time scales.

## II LITERATURE SURVEY

1. A Deep Learning Approach for Mapping Music Genres

In this article a CRNN structure on MagnaTagATune dataset is proposed. The AUC-ROC index for the proposed architecture is 0.893 which shows its superiority rather than traditional structures on the same database. The merging mechanism to obtain 50 tags from the whole 188 existing tags of this dataset and simple CRNN architecture designed for tag discovering are the main contribution of this paper. CRNN architecture for tag recognition in music application is proposed in the current article. The dataset utilize in this paper is MagnaTagATune which has 188 different tags and raw music in the mp3 format. To have better visibility on tag discovering the tags are reduced to 50 with merging similar features such as genre and instrument. The mp3 format is transfer to mel-spectrogram signal in pre-processing stage which has rich information around the music. The prepared information as a audio metric is fed to the proposed CRNN architecture which is designed to understand local and temporal features with inside CNN and RNN networks respectively. [1]

 

2. Convolution recurrent neural networks for music classification

Author introduces a convolution repeated neural network (CRNN) for music tagging. CRNNs cash in of convolution neural networks (CNNs) for native feature extraction and repeated neural networks for temporal report of the extracted options. Author has a tendency to compare CRNN with three CNN structures that are used for music tagging whereas dominant the quantity of parameters with reference to their performance and coaching time per sample. Overall, author have a tendency to found that CRNNs show a robust performance with reference to the quantity of parameter and coaching time, indicating the effectiveness of its hybrid structure in music feature extraction and have report. [2]

3. Deep Neural Networks: A Case Study for Music Genre Classification

In this paper we have a tendency to explore ascendible approaches to music classification by genre on a public dataset and examine the difficult aspects of this drawback. During this paper, author has a tendency to explore a two-layer neural network with manifold learning techniques for expressive style classification. we have a tendency to compare the classification accuracy rate of deep neural networks with a collection of well-known learning models together with support vector machines (SVM) there experimental results show that neural networks square measure comparable classic learning models once the information is depicted in an exceedingly wealthy feature house. [3]

4. Improved music feature learning with deep neural networks

In this paper author have a tendency to examine three ways that to enhance feature learning for audio information mistreatment neural networks: first using corrected Linear Units (ReLUs) rather than commonplace sigmoid units; second using a powerful regularization technique known as Dropout; third using Hessian-Free (HF) improvement to enhance coaching of sigmoid nets. Author has a tendency to found that the rectifier networks learnt higher options than the sigmoid networks. Author has a tendency to conjointly demonstrate the capability of the options to capture relevant data from audio information by applying them to genre classification on the ISMIR 2004 dataset. [4]

5. Long Short-term Memory Recurrent Neural Network based Segment Features for Music Genre Classification

In the typical frame feature primarily based expressive style classification strategies, the audio information is depicted by freelance frames and therefore the serial nature of audio is completely unheeded. If the serial data is well sculptured and combined, the classification performance is often considerably improved. The long STM (LSTM) repeated neural network (RNN) that uses a collection of special memory cells to model for long-range feature sequence has been with success used for several sequence labeling and sequence prediction tasks. During this paper, author has a tendency to propose the LSTM RNN primarily based section options for expressive style classification. The LSTM RNN is employed to find out the illustration of LSTM frame feature. The section options square measure the statistics of frame options in every section. [5]

6. Musical Genre Classification of Audio Signals

In this paper, the automated classification of audio signals into a hierarchy of musical genres is explored. A lot of specifically, three feature sets for representing timbral texture, metric content and pitch content square measure planned. The performance and relative importance of the planned options is investigated by coaching applied math pattern recognition classifiers mistreatment real-world audio collections. Each whole file and period of time frame-based classification schemes square measure delineated. Mistreatment the planned feature sets, classification of sixty one for 10 musical genres is achieved. This result's cherish results rumored for human music classification. [6]

## III Architecture of Deep Neural Network

A deep neural network (DNN) is an Artificial Neural Network with multiple hidden layers between the input and output layers. Similar to shallow ANNs, DNNs can model complex non-linear relationships. DNN architectures generate compositional models where the object is expressed as a layered composition of primitives. The other extra layers enable composition of features from lower layers. Deep architectures include many variants of a few basic approaches. Each architecture has found success in specific domains. It is not always possible to compare the performance of multiple architectures, unless they have been evaluated on the same data sets. DNNs are feed forward networks in which data flows from the input layer to the output layer without looping back. Recurrent neural networks (RNNs), in which data can flow in any direction, are used for applications such as language modeling long short-term memory is particularly effective for this use. [1]

## IV Architecture of CRNN (Convolutional Recurrent Neural Network)

Recurrent Neural Networks are in the family of feed-forward neural networks. They are different from other feed-forward networks in their ability to send information over time-steps. The CRNN consists of 4 parts: (1) at the highest of the design, a time frequency illustration of the information (a context window of F log band energies over T frames) is fed to $Lc \in N$ convolutional layers with non-overlapping pooling over frequency axis; (2) the feature maps of the last convolutional layer area unit stacked over the frequency axis and fed to $Lr \in N$ repeated layers; (3) one feed forward layer with sigmoid activation reads the ultimate repeated layer outputs and estimates event activity chances for every frame and (4) event activity chances area unit binarized by thresholding over a relentless to get event activity predictions. During this structure the convolutional layers act as feature extractors, the repeated layers integrate the extracted options over time so providing the context info, and eventually the feed forward layer turn out the activity chances for every category. The stack of convolutional, repeated and feed forward layers is trained conjointly through back propagation. [2]

Musical genres area unit labels created and utilized by humans for categorizing and describing the large universe of music. Musical genres haven't any strict definitions and bounds as they arise through a fancy interaction between the general public, marketing, historical, and cultural factors. This observation has light-emitting diode some researchers to counsel the definition of a brand new genre classification theme strictly for the needs of music info retrieval. But even with current musical genres, it's clear that the members of a specific genre share bound characteristics generally associated with the instrumentation, Sapphic structure, and pitch content of the music. Automatically extracting music info is gaining importance as the way to structure and organize the more and more giant numbers of music files obtainable digitally on the net. It's terribly probably that within the close to future all recorded music in human history is going to be obtainable on the net. Automatic music analysis are going to be one among the services that music content distribution vendors can use to draw in customers. Another indication of the increasing importance of digital music distribution is that the legal attention that firms like Napster has recently received. Genre hierarchies, generally created manually by human consultants, area unit presently one among the ways that wont to structure music content on the net. Automatic music classification will doubtless automatism this method and supplies a very important part for a whole music info retrieval system for audio signals. Additionally it provides a framework for developing and evaluating options for describing musical content. Such options are often used for similarity retrieval, classification, segmentation, and audio thumb nailing and type the muse of most planned audio analysis techniques for music. Disadvantage of existing system are:-

1. Manual Classification is done in existing system.

2. Method used in Existing system takes more amount of time in Classification and Labeling.

## V CONCLUSION

In this paper the survey on Music Genre Recognition is done. In this survey we have studied six papers on the base paper topic. According to this survey we can conclude that Music information retrieval could be a help to understand the context in audio signals and categorize them based on various feature such as genre, instrument, mood and etc. This classification makes the audience to have better visibility in selecting their favorite music. More accurate to retrieval the existed tags in music cause more chance to attract larger amount of audience with enough satisfaction to select their music intelligently. Nowadays precise tagging retrieval mechanisms are provided by deep feature extracting and learning methods. They attempt to understand almost all information inside music from local to temporal features. Combination of CNN and RNN networks is an architecture which has potential to explore the required information for tag classification. In this paper, we reviewed various methods, by far the best method that we have come across in CRNN.

## VI REFERENCES

1. Arjun Raj Rajanna, Kamelia Aryafar, Ali Shokoufandeh, "Deep Neural Networks: A Case Study for Music Genre Classification," IEEE Transactions on Parallel Distributed Systems, vol. 23, no. 6, pp. 794–805, 2015.

2. Keunwoo Choi, Gyorgy Fazekas, Mark Sandler, "Convolution recurrent neural networks for music classification ," ACM Computing Survey, vol. 39, no. 1, 2016.

3. Siddharth Sigtia, Simon Dixon, “Improved music feature learning with deep neural networks,” IEEE Trans. Parallel Distrib. Syst., vol. 18, no. 5, pp. 577–588, 2015.

4. Jia Dai1, Shan Liang1, Wei Xue1, Chongjia Ni2, Wenju Liu1, “Long Short-term Memory Recurrent Neural Network based Segment Features for Music Genre Classification , vol. 25, no. 9, pp. 2245–2254, 2015.

5. B George Tzanetakis, Student Member, IEEE, and Perry Cook, “Musical Genre Classification of Audio Signals,” IEEE Transactions on Parallel and Distributed Systems, vol. 17, no. 5, pp. 403–418, 2012.

6. Mehdi Roopaei, Paul Rad, Mo Jamshidi, Deep Learning Control for Complex and Large Scale Cloud Systems, Intelligent Automation and Soft Computing (AUTOSOFT), DOI: 10.1080/10798587.2017.1329245.3)

7. Li, Tom LH, Antoni B. Chan, and A. Chun. "Automatic musical pattern feature extraction using convolutional neural network." Proc. Int. Conf. Data Mining and Applications. 2010.

8. Nakashika, Toru, Christophe Garcia, and Tetsuya Takiguchi. "Local-feature-map Integration Using Convolutional Neural Networks for Music Genre Classification." Interspeech. 2012.

9. Sigtia, Siddharth, and Simon Dixon. "Improved music feature learning with deep neural networks." Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. IEEE, 2014.

10. Lai, Siwei, et al. "Recurrent Convolutional Neural Networks for Text Classification." AAAI. Vol. 333. 2015.

 