TEAM 2_Literature Survey

MACHINE INTELLIGENCE-UE20CS302

LITERATURE REVIEW

Name	SRN	Section
Ayush Singh	PES2UG20CS081	B
Ayushmaan Kaushik	PES2UG20CS082	B
Bhavini Madhuranath	PES2UG20CS088	B

Machine Learning and Deep Learning Techniques For Genre Classification of Bangla Music.

`2022 International Conference on Advancement in Electrical and Electronic Engineering`

Abstract

Music genre classification is extremely important for both music recommendation and acquisition of music data, as well as for music discovery. Despite the fact that Bangla music is extremely diverse in terms of its own style, there has been little notable work done to date to categorize Bangla genres using machine learning approaches.

Models are evaluated using f1-score, recall, accuracy and precision. As can be observed, the implemented deep neural network model was able to reach an accuracy of 77.68 percent. Index Terms—Bangla Music, Music Classification, SVM, RBF Kernel, Machine Learning, Deep Learning, Random Forest, KNN, SGD.

Methodologies

Preprocessing

- Label encoding is an essential pre-processing step in supervised learning for the ordered dataset. Each unique label is associated with an integer value in an ordinal encoding. This phase was achieved through the use of Sci-kit Learn's LabelEncoder, which encodes labels into numerical terms.

- In data analysis, feature scaling is used to standardize the range of independent variables or characteristics of data. It has been shown to dramatically increase the performance of machine learning algorithms. It also makes adjustments to data that has diverse scales in order to eliminate biases caused by large outliers.
- The curse of dimensionality relates to the concept that as.more input features are included in a computer modeling activity, it becomes. more difficult to assess the results. Principal Component Analysis (PCA) was used to reduce the dimensionality of the data.

Modeling

- There are lots of classification algorithms in machine learning. Also, deep neural networks are massively used for classification problems. In our approach, to classify the Bangla music dataset, we deployed five machine learning algorithms and a deep neural network.

Implementation and evaluation

-We use the loss function 'modified huber' as well as K-Nearest Neighbors and Stochastic Gradient Descent (SGD) to train our models.As a hyperparameter, we use 15 neighbors for KNN.

-We developed a sequential deep neural network for the music classification experiment. The first, second, third and fourth layer has 128, 64, 32 and 16 neurons, respectively.A total of 150 epochs are used to train the model, with a batch size of 16. Dropouts account for 0.4, 0.3, and 0.2 percent of the neurons for the corresponding layers, respectively.

Conclusion

- For each of the machine-learning models, we calculated recall, precision, f1-score, and accuracy. In comparison to Deep Learning, the deep neural network performed the best with an accuracy of 77.68%. Random forest, SVM-RBF, Linear SVM, SGD-SVM, KNN has marginal accuracy in comparison to other models.
- We propose a system that makes use of both machine learning algorithms and deep learning techniques to categorize Bangla music into six classes. Our models achieve a high level of accuracy, despite the fact that the dataset was not large enough. We intend to use more data to improve overall performance and automatize the procedure.

Music Genre Classification using Machine Learning.

`Fourth International Conference on Electronics, Communication and Aerospace Technology (ICECA-2020)`

Abstract

Music industry has undergone major changes and the ever-growing customer base has increased the market for different music styles. The manual ranking of music is a repetitive, lengthy task and the duty lies with the listener. This research work has compared few classification models and established a new model for CNN, which is better than previously proposed models.

Methodologies

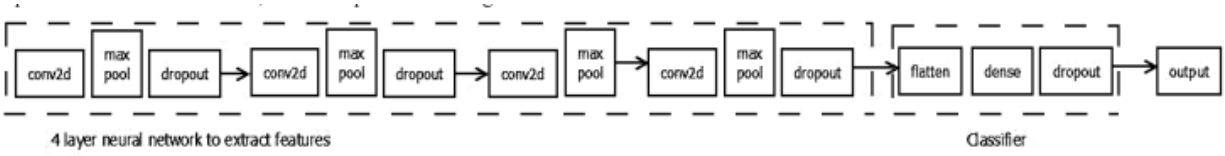
Dataset preparation

- There are two well-known datasets available for Music. Genre Classification, the FMA dataset [9] and the GTZAN dataset [10]. This paper has 10 classes, with each class containing 100 audio clips. Each clip is 30 seconds long in.wav format and samples at.22050Hz, 16bit

Preprocessing

- Each audio file is converted into itsmel-spectrogram then audio clips are loaded using the librosa library and generated. After the spectrogram was generated it was sliced into 64 strips. This increased our samples 64 times to a total of 64,000 samples each of dimension 480x10.
- Every audio signal can be represented in two ways:. 1) Assembling an audio signal, or 2) Feature Extraction - where audio features are extracted that are relevant for solving a problem. These definitions are inspired by the work of author in [11] and [12]..

Modeling



-Images are passed to a deep neural network consisting of two sub-networks. The first is a four-layer convolution neural network for extracting features from the images. These extracted features are then passed to the second sub-network for classification. In the end a dense layer is used to predict the audio quality.

Implementation and evaluation

- Our proposed models are implemented on a PC having TensorFlow 1.12.0, tflearn 0.3.2, Keras 2.2.4, OpenCV 3.4.3 and Python 3.5.2 using an Intel processor and 32 GB RAM. For measuring the performance of deep neural networks, the GTZAN dataset is used.
- The three parameters are used to estimate the performance of the model, sensitivity, specificity, and accuracy. Here sensitivity tells us how well our models classify a particular class, and specificity tells us how well our model is classified for non-current class. Accuracy tells us about the overall ratio of correctly detected events.

Conclusion

- The proposed research work has utilized the GTZAN dataset and produced multiple models to complete this task in this piece of music classification. The proposed model has used multiple inputs for various models along with the audio mel-spectrogram and transferred this to our CNN, and various sound file characteristics stored in the ANN, SVM, MLP and Decision Tree csv archives 91%, equivalent to the human understanding of genre with highest accurate achievement.Since, some styles were quite distinctive and some rather distinctive such as the country and the rock genre were confused with other styles, although traditional and blues were easily identified

Comparison of SVM, KNN, and NB Classifier for Genre Music Classification based on Metadata

`2 0 2 0 International Seminar on Application for Technology of Information and Communication (iSemantic)`

Abstract

Spotify is one of the most popular music streaming platforms in the world. This research aims to test and test the performance of music genre classification using three different classifiers. The SVM classifier has the best classification with 80% accuracy, then followed by KNN with 77.18% and NB with 76.08%.

Methodologies

- Data acquisition and selection: Take the Spotify music dataset using www.crowdai.org website. The dataset has a total of 228,159 music, consisting of 26 genres and each genre has 18 features. Each genre is taken 6000 sample samples and each feature is analysed over a period of several hours.
- Preprocessing: In the first stage, training and testing data were distributed, 80% of training data, and 20% of testing data.
- Feature Extraction: Next 18 metadata features are extracted except genre features, then 17 are converted to numeric values, and normalized for training on SVM. Feature selection is performed using the chi-square method as described in section 1.1 of the Open Data Platform (OCP) pre-training.
- Classification: at this stage, the training process is done first based on data and the selection of features that have been determined, then the testing process is done using each method
- Evaluation and measurement: the results of the testing process for each method are then measured for their accuracy

Implementation and evaluation

- This study uses the Spotify music dataset which has.228,159 music with 26 genres and 18 features. To reduce the computational time of 26 genres, 5,6,7, an 8 favorite genres were chosen. The data is separated into training data and testing data, where training data is 80% music for each genre. The SVM kernel used is a radial basis function (RBF) wherein this kernel two variables are needed, namely gamma (y) and C. Gamma value are used to determine the extent of the influence of each data with a borderline, while C is a hyperparameter which can determine the margin on the data used.
- SVM-RBF was chosen because it is considered quite reliable and popular to use but requires a relatively shorter running time than KNN and NB. In the KNN classifier, no single k value is used, but k values between 1 and 100 are sought.
- The classification of music genres produced based on metadata can produce relatively the same performance as the Audio feature extraction approach, even though the KNN and NB classifiers appear to be superior.

Conclusion

- SVM-RBF proved to be superior in terms of accuracy and computational time to KNN and NB for music genre classification based on metadata features, compared to the other two methods. This research is applied to the world's leading streaming platforms such as Spotify, Pandora and Rdio.

Music Genre Recognition Using Spectrograms

`2 0 2 0 International Seminar on Application for Technology of Information and Communication (iSemantic)`

Abstract

In this paper we present an alternative approach for music genre classification that converts the audio signal into spectrograms and then extracts features from this visual representation. On a very challenging dataset of 900 music pieces, we have demonstrated that the classifier trained with texture compares similarly to the literature.

Methodologies

- In this work we have used the idea. of time decomposition [9] in which an audio signal S is. decomposed into n different sub-signals. We have used three 30-second segments from the beginning (Ybeg), middle (Ymid), and end (Yend) parts of the original music. After generating the spectrograms, the next step is to extract the features from the images. This involves using the well-known GLCM texture textures as features.
- The best setup we have found is d = 1 and angles d and θ = 0, 45,. Figure 2 depicts the zoning mechanism used for this work.
- The spectrograms (Ybeg, Ymid , and Yend) were trained using the Support Vector Machine (SVM) classifier introduced by Vapnik in [10]. Normalization was performed by

linearly scaling each attribute to the range [-1,+1]. Different parameters and kernels for the SVM were tried out but the best results were yielded using a Gaussian kernel. By zoning the images we can extract local information and try to highlight the specificities of each music. The rationale behind the zoning and voting scheme is that signals may include similar instruments and similar patterns which leads to similar areas in the spectrogram images.

Implementation and evaluation

- The Sum rule explores the diversity produced by the two different classifiers, and achieves an recognition rate of 67.2%. That is an improvement of 7 percentage points compared with the baseline system presented in Table I. The results compare well to those published in the last MIREX contest on audio genre classification.

Conclusion

In this paper we have presented an alternative approach to music genre classification which is based on texture images rather than audio signals. Using the Max rule we were able to reach an improvement of about 7 percentage points in the recognition rate. Our future works will be focused towards the develop- ment and tests of other texture features and other strategies for zoning.

Music Genre Classification and Recommendation by Using Machine Learning Techniques

Abstract

Digital signal processing methods were used in this study to extract the acoustic characteristics of the music, and machine learning techniques were used to classify the music's genres and generate music recommendations. Convolutional neural networks, another deep learning technique, was also employed for genre categorization, music suggestion, and performance evaluation of the resulting data. The GTZAN dataset was employed in the study, and the SVM algorithm had the most success overall.

Methodologies

- Dataset: GTZAN
- Feature Extraction:
- Zero Crossing Rate
- Spectral Centroid
- Spectral Contrast
- Spectral Bandwidth
- Spectral Rolloff
- Mel Frequency Coefficient of Cepstrum-MFCC

- Machine learning techniques:
- KNN
- Naive Bayes
- Decision trees
- SVM
- Random Forest
- CNN

Implementation and evaluation

Raw Audio: After taking a 30 second chunk, audio data is immediately transmitted to the network. Since audio data is a one-dimensional vector, 1D convolution layers are utilised rather than 2D convolution layers.

Short Time Fourier Transform (STFT): Audio data is subjected to Short Time Fourier Transformation before being sent to the network. This changes the frequency domain of a one-dimensional time series into a two-dimensional one. The window size and hop length parameters for this transformation are both set to 2048.

Mel Frequency Channel Cepstrum: Based on a mel scaled spectrogram, cepstrum is a representation of the short-term power spectrum of a sound. MFC Coefficients make up the MFC.

A graphical user interface has been developed in Python to compare all results in this part, including those from deep learning and conventional acoustic characteristics. from unprocessed music, extract acoustic features Both the Librosa library and Keras , which are used for deep learning partition, are helpful libraries. Through this interface, a user can load raw audio and extract features. Additionally, utilising the interface, classifier meta-parameters like the kernel type of SVM and the k value of KNN as well as some feature extraction method parameters like window size, window type, and overlap ratio may be changed.

Conclusion

The goal of this project is to categorise and recommend songs based on auditory data that were obtained using convolutional neural networks and digital signal processing techniques. The study was done in two stages: figuring out how to get the features that would be utilised in recommendations and creating a service that suggests music based on user requests. After performing feature extraction using digital signal processing techniques, CNN was trained to serve as a backup feature extraction approach. The optimal categorization method and recommendation outcomes are then determined by using the acoustic characteristics of the music. The findings listed in the previous tables show that SVM outperformed other methods in terms of classification outcomes. Additionally, very minor performance modifications were produced by altering the window type and size.

Automatic Music Genre Classification using Convolution Neural Network

Abstract

The classification of music by genre is crucial in the modern world since the number of music tracks, both online and offline, is growing quickly.

We must appropriately index them in order to have greater access to them. To retrieve music from a vast collection, automatic music genre classification is crucial. The majority of the existing methods for categorising music genres rely on machine learning. We give a music dataset with ten distinct genres in this article. The system is trained and classified using a Deep Learning technique. Convolution neural networks are employed in this instance for training and classification. For audio analysis, feature extraction is the most important step. For sound samples, the Mel Frequency Cepstral Coefficient (MFCC) is employed as a feature vector. The suggested system categorises music based on this.

Methodologies

The music database is divided into many genres using the proposed system. The system can be trained using a variety of databases. According to the literature, GTZAN and Million Song Dataset (MSD) are the most often utilised databases. Here, we intend to employ the MSD in our strategy. This is a freely downloadable collection of songs that spans several genres. The collection is made up of 280 GB worth of audio tracks. The process of removing elements from this enormous collection of music is incredibly laborious.

Initially the database of the music is created. Then each song has to go through a preprocessing stage. The librosa library in Python is used for Feature Vector Extraction. Use this particular programme only for audio analysis.

Every audio file is taken, and the feature vector is then retrieved from there. MFCC refers to the feature vector that was extracted. By recording the approximate form of the log-power spectrum on the Melfrequency scale, the MFCCs encode the timbral characteristics of the music signal.

Implementation and evaluation

A convolutional neural network (CNN) is made up of one or more fully connected layers after one or more convolutional layers, just like a typical multilayer neural network. Each neuron takes input from the feature vectors, which are then dot products with the weights and sent to the subsequent layers, where a non-linearity may or may not follow.

The Anaconda package for Python was used for the evaluation. The deep learning package utilised was Tensorflow. The technique was implemented using an Intel Xeon CPU E5-2630 v4 with 2.20GHz, 10 Core(s), and 20 Logical Processors, together with 32GB of RAM. The

number of iterations was increased from 10,000 to 100,000 in increments of 10,000. The results of the learning accuracy tests using the Mel Spec and MFCC feature vectors were 76% and 47%, respectively, as shown in Fig 5. Mel Spec requires more time to learn whereas MFCC needs less time to converge. Prediction is quicker once the model has been built.

Conclusion

A method for automatically classifying music genres based on Convolution Neural Networks is presented in this paper. Mel spectrum and MLCC are used to calculate the feature vectors. The python-based librosa software aids in feature extraction and hence assists in supplying useful parameters for network training. For the Mel Spec and MFCC feature vectors, the learning accuracies are proven to be 76% and 47%, respectively. Therefore, this system holds promise for categorising a large database of music into the appropriate category.

The future development will concentrate on improving the method to categorise the songs according to mood. This will be useful in determining the genre of music that can make someone feel less stressed while listening to it.

Music Genre Classification using Machine Learning

Abstract

Both the music produced in recent years and the music industry as a whole have undergone significant changes. The market for various music styles has expanded along with the company's steadily expanding consumer base. Not only can music unite people, but it also sheds light on other civilizations. Therefore, it is crucial to categorise music according to genres in order to suit people's needs. It is the responsibility of the listener to manually rank the music because it is a tedious and time-consuming operation. The proposed study work analysed a number of classification models and created a new model for CNN that is superior to earlier suggested methods.

Music genre classification has been the subject of extensive investigation. Based on the type of dataset employed, these studies can be broadly divided into two groups. There are two well-known datasets: the GTZAN and FMA datasets. CNN is a neural network with a topology resembling a grid. This grid may be linear, similar to time series data, or 2D, similar to an image.

CNN uses a technology that lowers the processing demands, resembling a multilayer perceptron. In addition to CNN, other methods for comparative analysis include support vector machines, artificial neural networks, multilayer perceptrons, and decision trees.

Methodologies

Dataset Preperation

The FMA dataset contains information about the audio properties of 8000 different songs from 8 different genres, and the GTZAN dataset contains 1000 audio files from 10 different genres. These two well-known datasets are available for music genre classification.

The GTZAN dataset, is also employed in this study. Blues, Classical, Country, Disco, Hip-Hop, Jazz, Pop, Metal, Reggae, and Rock are among the ten categories in this dataset. Each class has 100 audio clips, each of which is in.wav format and has a duration of 30 seconds with sample rates of 22050Hz and 16bit.

Data Pre-Processing

This dataset contains 100 audio clips for one class which is not enough to get a good accuracy. Either one could use a dataset with more audio clips or pre-process the dataset in such a way that it increases the number of training and testing samples.

Spectogram Generation: Each audio file is converted into itsmel-spectrogram then audio clips are loaded using the librosa library and generated the mel-spectrogram for each audio clip. After the spectrogram was generated it was sliced into 64 strips. This increased our samples 64 times. We observed a total of 64000 samples each of dimension 480x10 for training and testing. So, we divided out 64000 images into for 44200 testing, 7000 for validation and 12800 for testing. Fig. 1 shows the original image and the strips generated after processing.

Implementation and evaluation

Feature Extraction: Every audio can be represented in form of an audio signal and this signal has different features. The audio features are extracted that are relevant for solving the problem. These features are divided into 2 sub categories.

Convolution Neural Network: As we can see, even in an image of 480x10 pixels, there are traits and features that are unique to each class. These photographs serve as input to our CNN model.

CNN Model: The training images are passed i.e. the sliced images of the spectogram to our deep neural network for comprising of two sub-networks. The first neural network is a four-layer convolution neural network for extracting features from the images. These extracted features are then passed to the second sub-network for classification. This network is fully connected network containing two fully connected layers. In the end a dense layer is used to predict the genre of the audio.

Layers used in CNN:

- Convolution
- Max Pooling
- Dropout

Other models used

Artificial Neural Network: This computing model is based on how the human brain functions. This is the case because of how the information moves through the brain. The neural network's structure is impacted as information passes through its neurons, allowing it to learn from input and output. When a complex relationship between the input and the output needs to be discovered, ANN, a nonlinear data model, is used.

MLP: This kind of logistic regressor involves the insertion of a hidden or intermediary layer.

This layer contains nonlinear activation functions (tanh or sigmoid). By adding as many hidden levels as the user requires, one can deepen the architecture.

Decision Tree: This classifier is employed in the classification of many classes. A binary tree can be used to represent this kind of paradigm. A sequence of questions about the dataset (related to its features/attributes) is presented starting from the root node to all internal nodes, and the nodes are then further divided into new nodes with distinct properties. The classes into which the dataset is divided are represented by the leaves of the tree.

Conclusion

Sensitivity, specificity, and accuracy of the model are estimated using the three parameters. Here, sensitivity and specificity describe how well our models categorise a specific class and a non-current class, respectively. The entire ratio of accurately identified occurrences is what accuracy reveals. Here, accuracy for the overall findings and sensitivity and specificity for each class are calculated. TP for a class, let's say X, is all the instances of that class that are used to calculate sensitivity and specificity. All non-X instances that are not classed as X are referred to as TN, whereas all non-X instances that are classified as X are referred to as FP.