International Journal of Scientific & Engineering Research Volume 13, Issue 12, December-2022
ISSN 2229-5518

Lyric Based Multilingual Music Genre Classification

K. Jhansi Lakshmi Bai, Kallam Lalithendar Reddy, Boppana Shashank, Gottipati Aditya, V. Hareen Sai

Department of computer science & Engineering, VNRVJIET

Abstract— Lyrics based Multilingual Genre Classification still remains an interesting problem in the field of Music Information Retrieval (MIR). If we humans were to manually classify the genre, then it would be tedious task to classify each song. There often is a debate where a song fits, since boundaries are not clearly defined and genres are overlapping. Different people may think different genres, or even multiple genres to the same song. With the emergence of big data, Due to the overwhelming number of music resources that individuals possess, it is challenging to store and retrieve music data. To address the issue, music classification is developed to help people swiftly discover music genre. Modern methods involve classifying genre using acoustic features as well as feeding lyrics into neural networks. Some of these approaches fail to take into account word order which we think can be really important. This approach investigates the likelihood of identifying the genre of a song from its lyrics. We use Long Short Term Memory (LSTM) for classifying the genres based on the lyrics. Previously LSTM models are used for sentiment analysis, semantic analysis, and text classification. As lyrics are also text data, we would like to check the potential of LSTM in classifying Song lyrics into genres.

Index Terms— Information Retrieval, Language detection and Translation, Long Short Term Memory (LSTM), Text classification, Text preprocessing.

1 Introduction

The aim is to classify songs into genres of any language by using machine learning algorithm. In Fig. 1 these are some examples of genres into which we can classify the songs. In the modern days there are many songs that are being released every day in multiple languages. So classifying these many songs manually will take a long time and is very difficult task for human beings. There are many studies about the classification of the songs based on audio features into genres using machine learning algorithms. There are also no of studies that also looked into the lyric features to replace or assist audio features and there are studies that have only worked with a single language data sets, which would be unrealistic in the real world. This paper aims to extend these studies with multilingual data by using effective machine learning algorithm with regard to the classification of the songs written in different languages using lyric features. Characters create words, words create sentences, and sentences put together create song lyrics. The secret long-term relationships must be revealed. A specific type of recurrent neural network that can manage long-term dependencies is the LSTM. The work involves categorizing a supplied text into pre-determined categories. The use of LSTM for music genre categorization is examined in this work. The LSTM model outperforms earlier RNN and attention techniques while being comparable quicker. Since the usual habit is to recall long-term information, Long Short Term Memory (LSTM) aids in the "long-term dependence" need (or resolves the short-term memory difficulty).

Fig. 1 Instruments and Genres in music

2 Related work

The Text mining is frequently used to process natural English language to draw broad conclusions from a volume of text. The fundamental objective of this study is to develop models that classify a song's genre dependent on its lyrics. These classifiers must be able to interpret song lyrics and categorize them as belonging to one of the following musical genres: rock, country, jazz, pop and other. To categorize each song, this work uses classifiers like J48, Random Forest, KNN, and Naive Bayes algorithms. In order to ensure accurate results, especially when dealing with string values, the necessity of collection of data and pre-processing is also highlighted [1].

The term "music genres" refers to the categorization of music based on the interactions between artists, commercial pressures, and cultural factors. By highlighting commonalities between musical works or performers, they aid in the organization of music into collections. Automatic genre classification is challenging because it might be challenging to tell one genre from another and because genres sometimes overlap and have ill-defined borders. In this essay, we attempt to categorize a collection of songs available on Spotify using their lyrics into the primarily four genres. On the preprocessed data, we apply two Word Embedding algorithms, Word2Vec and Word2Vec with TF-IDF, to turn the words in the songs into vectors of real numbers. We used Deep Neural Networks, Support Vector Machine, Random Forest, XGBoost, and other machine learning techniques on the produced word vectors to predict the genre. In Simple Word2Vec and Word2Vec with TF-IDF, we were able to get mean accuracy of 61.70% and 71.05%, respectively, with the maximum accuracy of 65.0% and 74.0% using a 3 Layer Deep Learning model for both techniques [11].

In this study, instead of using the traditional method of audio feature analysis, we propose to build a system that analyses the dataset of song lyrics based on the features derived from the training stage and can predict the genre of the song that is presented to the model in the validation stage. For the aim of validation, the suggested approach takes into account five moods, each of which contains 100 songs. Based on an examination of the lyric language, the algorithm can determine the song's mood [9].

It has been demonstrated that learning global language representations is facilitated by using models like BERT and Distil BERT. In several language understanding challenges; these language representation models have shown astounding results. In this work, we use song lyrics to apply language representation models for categorizing musical genres. We compare the outcomes and computation times of language representation models and conventional deep learning models for music genre classification. BERT performs better than other models on single class and multi-class classification, according to experimental findings, with accuracy rates of 71.29% and 77.63%, respectively. However, when epoch length is taken into account, BERT outperforms DistilBERT by a factor of 4 [3].

To categorize a sizable collection of whole music lyrics. As lyrics have a hierarchical layer structure wherein words combine to generate lines, lines form segments, and segments make up a full song, we employ a hierarchical attention network (HAN) to exploit these levels and additionally learn the importance of the words, lines, and segments. According to experimental findings, the HAN performs better than both non-neural models and simple neural models, categorizing over a wider range of genres than earlier studies. We can also see via the learning process which lyrics or verses of a song the model considers crucial for categorizing the genre [13].

Numerous aural, grammatical, and metadata elements have been used to study the categorization of musical genres. For the goal of predicting the music genre of lyrics, classification using linguistic data often yields worse accuracy than classifiers created with audio information. In this work, we manually create criteria including rhyme density, readability, and the usage of profanity that were not included in earlier lyrical classifiers. In the end, the lyrics' music genre could be predicted with an accuracy of 0.56136 over all nine genres, which is an improvement over earlier research' results [12].

3 Proposed Methodology

Here is the process involved in classification of song genre based on lyrics using following phases:
1) Data Collection Phase.
2) Data Cleaning Phase.
3) LSTM Model.

3.1 Data Collection Phase

In the process of lyric-based classification of music genre the initial step is data collection. Song genre classification using lyrics is like working on a low-dimensional data. The dataset is taken from Kaggle named Multi-Lingual Lyrics for Genre Classification. There are four sources that make up the data. The original information was sent from the 2018 Textract Hackathon at Sparktech. This was improved with information from three more Kaggle datasets: 50,000 Lyrics Song lyrics from AZLyrics, dataset lyrics, and Spotify Valence are tagged. The other datasets lacked a Genre attribute other from the original Sparktech data. The team created a labelling mechanism using the spotify library, which leverages the Spotify API to obtain the genre of an artist, to address the absence of Genre tagging. Please be aware that the Spotify API gives a list of genres for each artist, thus they determined that the dominant genre for that artist was the one that was most frequently returned. In Fig.2. we can see that the raw data contains 290,155 song lyrics which includes 13 genres in 33 different languages. The final data set after removing null values and genre with less than 1000 song lyrics contains data of 249,289 song lyrics containing 10 genres in 33 languages. We take only five languages data which have more than 500 song lyrics.

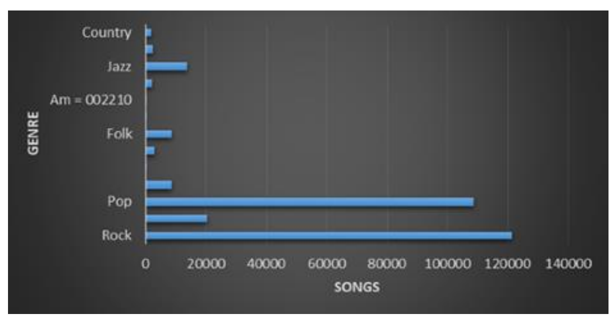

Fig. 2. Original Data Set

3.2 Data Cleaning Phase

The Data set taken contained lakhs of raw lyrics where many of them where not in the right format so We must first prepare and clean up our text data. Because it's unclear exactly what kind of preparation and cleaning we need to perform on text data, preparing text data presents a few additional difficulties than dealing with more conventional data types. The sole aspect of our data collection that we are concerned in at the beginning of our work is the column containing the complete text of every song lyric. Corpora in its natural form starts with just 1 dimension. As a result, we must decide how to preprocess our data and retrieve text document properties that we may utilize to train Machine learning and deep learning models later. We started by looking for non-values and removing songs containing NaN lyrics and songs that were simply musical notes. Then, after examining value counts for each genre, we made the decision to exclude Folk, Indie, and Also other as the first two had insufficient data and "Other" had no predictive value for our final classification job. Eight fundamental genres remained after all this cleaning: Rock, Pop, Hip-Hop, Metal, Country, Jazz, Electronic, and RB. We will be attempting to anticipate these target classes. Feature Engineering and Model Optimization: We accomplished the following using a combination of Pandas, Regex, Language Translation, Language Detection, and NLTK techniques: [4]

- Language Detection of Multi Lingual Lyrics.
- Now, translate the other language lyrics to a Target language (English).
- Remove the verse, chorus and Bridge words from lyrics using regular expressions. The Bridge words are present inside [], () or \\ using regular expressions we can substitute them with empty string.
- Encode all the text data to lowercase.
- Removing URL's.
- Removing HTML tags.
- So, the idea is to build a regular expression which can find all characters "< >" as a first incidence in a text, and after, using the *sub* function, we can replace all text between those symbols with an empty string.
- Removing punctuations, white space, special characters etc.
- Removal of Stopwords using Snowball Stemmer.

3.3 LSTM Model

Short-term memory is a problem for traditional neural networks. The vanishing gradient issue is also another significant negative. (While backpropagation, the gradient shrinks to the point where it tends toward zero.) By retaining the crucial information and identifying patterns, LSTMs significantly increase performance. The fact that LSTM works well for helping people remember crucial information is one incentive to employ it.

When predicting the class, alternative non-neural network classification algorithms would provide the output according to statistics rather than meaning since they are trained on many words as distinct inputs that are just words with no true meaning as a sentence. That implies that each word is assigned to a specific category. In LSTM, this is different. A multiple word string can be used in LSTM to determine the class into which it fits. When working with NLP, this is quite beneficial. The model will be able to determine the true meaning in the input text and will provide the most correct output class.

Because lyrics are fundamentally sequential in nature, it makes sense to utilize the Long Short-Term Memory (LSTM) model to assess how similar two lyrics are over time based on similarities between their sequences. In Fig.3 we can see the architecture of how LSTM model works on lyrics data. As the data set contain multiple language lyrics we need to build a model in such a way that the model understands the lyric language directly so for this task a package named langdetect is imported and use the detect function to detect the language of the particular lyric. Another packaged named google trans is imported and the Translator function is used to convert the lyrics of the song to a target language of English [2]. So, all the songs of different languages now converted to English where this becomes easier for the model to predict. Define two lists containing articles and labels. According to the parameter we previously defined, 90% for training and 10% for validation, we divided them into training set and validation set. Text data must first be converted to word sequence.

Tokenizer handles all the labor-intensive work for us. When an unknown word is encountered, oov token is used to insert a specific value. Fit on text will search the entire text and generate a dictionary, thus we want to utilize <OOV> for terms that are not included in the word index. By converting each text into either a series of integers or a vector with a binary coefficient for each token dependent on word count, the tokenizer class enables the vectorization of text corpora. internal vocabulary according to a collection of texts. The Pad Sequences method turns a list of sequences into a shape-based 2D Numpy array [6]. If a maxlen parameter is given, num timesteps is either that value or the length of the largest sequence in the list.

We start with an embedding layer and construct a tf.keras.Sequential model. One vector each word is stored in an embedding layer. Each word is represented as a 32-length vector in the embedded layer, which is the top layer. The LSTM layer, which has 64 memory units, is the following layer (smart neurons). When invoked, it transforms word index sequences into vector sequences. Words with comparable meanings generally have similar vectors after training. Since sigmoid and tanh functions are excellent substitutes for one another, we utilize them in lieu of one another. To create predictions for more than two classes in the task, you will use a Dense output layer and a sigmoid activation function because this is a classification problem. We include a 10-unit dense layer with sigmoid activation. Sigmoid transforms output layers into a probability distribution when we have numerous outputs. Because it is a multi class classification issue, log loss (or sparse categorical crossentropy in Keras) is employed as the loss function. The music genre is determined from the input by looking at the lyrics.

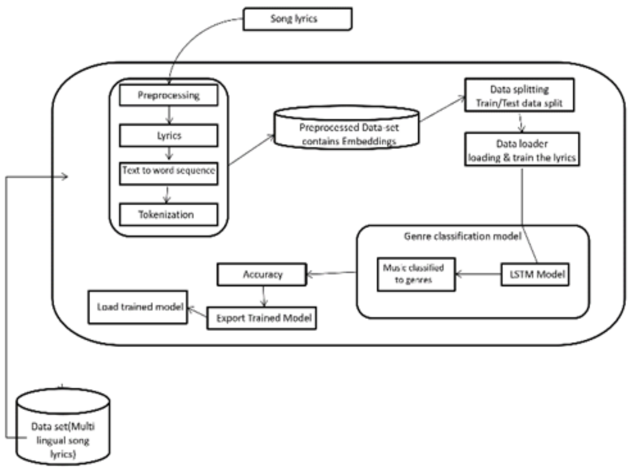

Fig. 3. LSTM Model

4 RESULTS

Before the usage of advanced models like NLP baseline models are created by taking a dummy dataset where the metrics like Precision, Recall and F1-score are measured. There are two baseline models the first one is the model that predicts the genre considering the factor of most-frequent genre even a biased data set could change the result and the second model predicts any random genre out of 10.

4.1 Baseline Models

Calculation of the baseline metrics is done. So, that we build a model that beats the baseline result. The Table 1. And Table 2. shows the baseline results of most frequent and random genre. Here are the formulae for calculating precision, recall and F1-score:

$$precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$F1 = 2 * \frac{precision * recall}{precision + recall}$$

Table 1. Most Frequent Genre

Metrics	Macro Avg	Weighted Avg
Precision	0.04	0.18
Recall	0.10	0.42
F1-score	0.06	0.25

Table 2. Random Genre

Metrics	Macro Avg	Weighted Avg
Precision	0.10	0.32
Recall	0.09	0.10
F1-score	0.06	0.14

4.2 LSTM Model

After applying the LSTM model to the dataset, we found different accuracies with different numbers of word length, epochs and batch size, if the page size would be too much large then model is mis behaving with the result. As genre should be classified to around 5 different languages the accuracy differs with respect to the language of the lyric although we converted it into the English lyrics. After processing the English language, it gives an accuracy of about 51%. The Italian lyrics was a bit different than others where around 2000 song lyrics were trained and the accuracy is 87% which is the topmost among the other languages, the model has loss of 0.59 and accuracy about 0.8784. When the Spanish lyrics were processed, we got an accuracy of 75% with a loss of 0.64% where model is working quite well when tested with own data of new songs and Portuguese language lyrics has an accuracy of 62.3% and this is obtained which has loss of 0.94. And the Indonesian lyrics got an accuracy of 76% with a loss of 0.94. The Table 3. Shows the loss and accuracy results of an LSTM model on different languages. The figures Fig. 4, Fig. 5, Fig. 6, Fig. 7 and Fig. 8 visualizes the training and validation accuracy of English, Indonesian, Italian, Spanish and Portuguese languages. Adam optimizer was used to all lyrics. In deep learning, optimizers are used to adjust a neural network's parameters in order to reduce the cost function. Therefore, selecting the right optimizer is crucial and can make the difference between a successful training and a terrible training. Most of the time, Adam is the most effective adaptive optimizer. It works well with sparse data. Overall the model is predicting with a good accuracy on an average of around 70.55%.

Table 3. Accuracy results of the Model

Language	Loss	Accuracy
Italian	0.59	87.8
English	1.46	51.1
Portuguese	0.94	62.3
Spanish	0.64	75.0
Indonesian	0.94	76.5

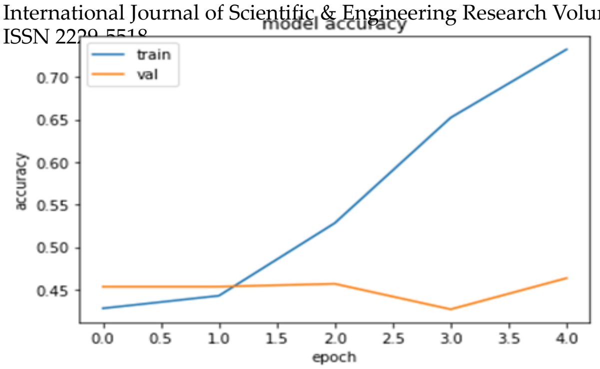

Fig. 4. Training and Validation accuracy of English

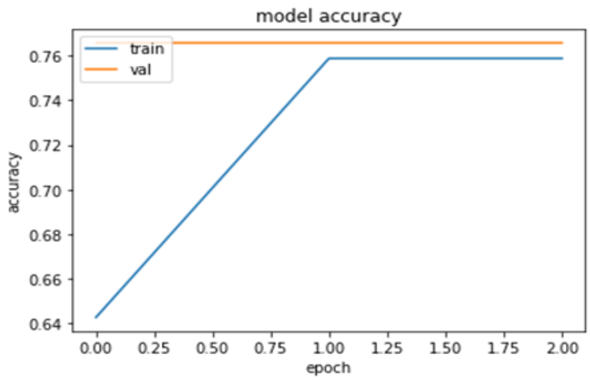

Fig. 5. Training and Validation accuracy of Indonesian

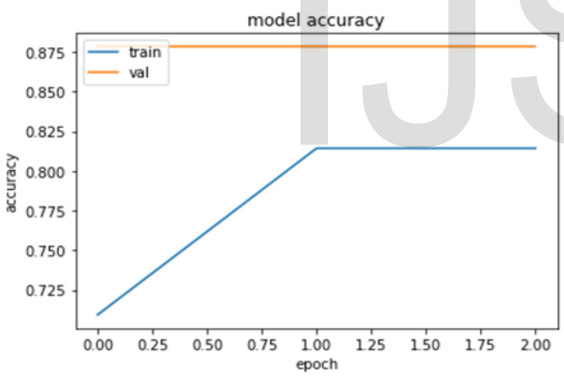

Fig. 6. Training and Validation accuracy of Italian

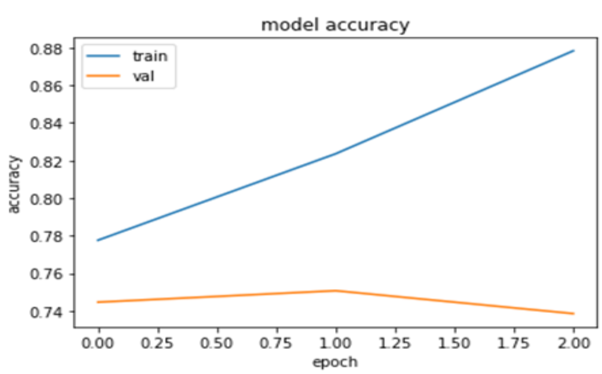

Fig. 7. Training and Validation accuracy of Spanish

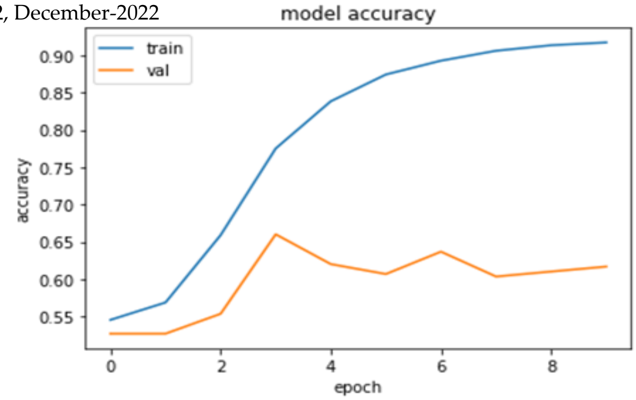

Fig. 8. Training and Validation accuracy of Portuguese

5 Conclusion

The Music Genre Classification is very important for companies like Spotify, Wynk , Gaana etc; to know about the user preferences of what type of song the user likes the most and wills to listen to them. So, that the system suggests the user different songs in his next visit in the same genre. Generally, Genre classification task is done based on acoustic signals i.e, audio. Using audio for this task makes it slow and need huge storage to store the data of millions of songs.

We use an LSTM Model to classify music genre based on their lyrics. When we require our model to take into account long-term dependencies, LSTM performs better than the other models. The LSTM outperforms RNNs due to its capacity to forget, recall, and update the information.

References

[1] Canicatti., A. (2016). Song Genre Classification via Lyric Text Mining. International Conference on Data Science(ICDATA).

[2] David, D. (2021, 1 20). How to Detect and Translate Languages for NLP Project. Retrieved from www.towardsdatascience.com: https://towardsdatascience.com/how-to-detect-and-translate-languages-for-nlp-project-dfd52af0c3b5

[3] H, A., E, F. C., Bolucu, Y. S., & B, C. (2021). Language representation models for music genre classification using lyrics. International Symposium on Electrical, Electronics and Information Engineering.

[4] Harshith. (2019, 11 21). Text Preprocessing in Natural Language Processing. Retrieved from www.towardsdatascience.com: https://towardsdatascience.com/text-preprocessing-in-natural-language-processing-using-python-6113ff5decd8

[5] Laurier, & Grivolla, C. (2008). Multimodal music mood classification using audio and lyrics. International Conference on machine learning and appllications(IEEE).

[6] Li, S. (2019, 12 8). Multi Class Text Classification with LSTM using TensorFlow 2.0. Retrieved from www.towardsdatascience.com: https://towardsdatascience.com/multi-class-text-classification-with-lstm-using-tensorflow-2-0-d88627c10a35

[7] Lima, A., & Sousa, R. D. (2020). Brazilian lyrics based music genre classification using a BiLSTM network. International Conference on Artificial Intelligence and Soft Computing.

[8] MALATO, G. (2022, 7 2). Does your model beat the baseline? Retrieved from www.yourdatateacher.com: https://www.yourdatateacher.com/2022/07/04/does-your-model-beat-the-baseline/

[9] Meenakshi, K. (2020). Music Genre Classification using Lyric Mining based on tf-idf. International Journal of Recent Technology and Engineering (IJRTE).

[10] Pereria, & Costa, R. M. (2019). Representation Learning vs. Hand crafted features for music genre classification. International Joint conference on neural networks(IJCNN).

[11] Rajpal, K. A., & D., R. (2018). Genre classification using word embeddings and deep learning. International Conference on Advances in Computing, Communications and Informatics (ICACCI).

[12] Thompson, C. (2021). Lyric-Based Classification of Music Genres Using Hand-Crafted Features. International Journal of Undergraduate Research.

[13] Tsaptsinos, A. (2017). Lyrics-based music genre classification using a hierarchical attention network. INTERNATIONAL CONFERENCE ON MULTIMEDIA AND EXPO.

[14] Ying, T. C. (2015). Lyric based genre classification using variant tf-idf weighting schemes. Journal of Applied sciences.

[15] Ying, T. C., Doraisamy, S., & Nurliyana, L. (2012). Genre and mood classification using lyric features. Internation conference on information retrieval and knowledge management.M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.