**Song Lyrics Genre Detection Using RNN**

**ABSTRACT**

Digitalization of music is the new trend, and preferences of individuals are highly rated. Millions of songs are being streamed in the music applications. The companies providing these services need to sort and arrange a wide range of music tastes for all of its users. On top of that, fresh music from various artists in a wide spectrum of genres are popping up every day. To keep track of all this, a classification system can be handy. So, we propose an RNN based model based on Natural Language processing to classify the songs based on their lyrics into different genres. Additionally, this tool can be handy to the music lovers for quickly identifying which genre a particular song belongs to. In this paper, we apply Long Short Term Memory (LSTM) model with both Universal Serial Embedder (USE) and Bert embedders. A comparative study is performed to understand which combination of models works based to classify the genres based on lyrics. From our results, on the basis of accuracy of the model, we found that USE embedder with LSTM gives a slightly better performance than Bert embedder. The LSTM model with USE embedding gave the highest accuracy of 83.42% when trained over a range of five folds.

1. **INTRODUCTION**

Digitization of songs took the music industry by a storm. More and more artists are releasing fresh music in digital marketing rather than physical means. This enables the artist’s music to spread all over the world. Services such as Amazon, Apple music, Spotify, etc. are providing consumers services for direct download and continuous streaming of their favorite music. Maintaining this huge database has become a prominent issue for the digital music providers as well as third party entities. The data picked for an individual song can include the name of the song, artist’s name, acoustics, genre, bass measure, tempo, etc. The websites which display the music lyrics for a song are developing APIs to access massive amounts of lyrical data to the developers.

Genre in music is very diverse. Pop culture has segregated several genres depending on the topic it discusses. So mainly, the lyrical part of song has the potential to decide the genre. For classification, we can use CNN, Naïve Bayes or RNN. But natural language process usually gives a higher accuracy as compared with traditional models. So, to build this model, we use RNN Long Short Term Memory(LSTM) model for prediction of genre based on lyrics. On top of this, we are going to use two separate embedders, which are Universal Serial Embedder (USE) as well as Bidirectional Encoder (BERT) embedder.

We perform a comparative study with these embedding techniques to find the best performing model for real life deployment. This analysis can be helpful for the developers of music websites, but also to the producers and musicians, to determine the characteristics of the lyrics and to understand which genre their song will typically fall into. Additionally, music enthusiasts can use this model to predict the genres of their favorite music.

**2. LITERATURE REVIEW**

**2.1 Related Work**

[1] This paper compares the language representation models with traditional deep learning models. They used BERT, DistilBERT and RNN Model BiLSTM with single label and multi label classification. They found out that BERT performs well with an accuracy of 77.63% for one label, and 71.29% for multi label classification. [2] This paper discusses songs being classified into four genres of Christian, Metal, Country and Rap. They used 2 processing techniques of TFIDF and BOW. Classification is done using CNN, SVM, Naïve Bayes, XGBoost. They found that TFIDF outperforms BOW. [3] They implemented Transformer classifier on audio clips of 1000 songs in 10 different Genres. Each audio clip is further split into smaller pieces. They used 2 methods highest incidence and respective probability vector. Achieved highest accuracy of 76%. [4] In this paper, Brazilian songs are classified into 14 genres using models like SVM, Random forest and Bi-LSTM. The embedding techniques used are Word2Vec, Wang2Vec, FastText and Glove. They observed that the BLSTM Network with Wang2vec word embedding provided the best F1-Score result. [5] In this paper, song lyrics is classified into 4 genres. Embedding is done using Word2vec and Word2Vec with TFIDF. Classification is done with SVM, Random Forest, XGBoost, Deep Neural Networks. Overall, Deep Neural Network with 3 layers and Word2Vec TFIDF model outperforms the other models with an accuracy of 74%.

[6] They examined different approaches to classify the genre using lyrics. They used 4 datasets to classify into 7 genres. Pre-processing techniques are applied, embedding is done using Bag of Words. Classifiers like Naïve Bayes, Support Vector Machine, XGBoost, Random Forest, LSTM and CNN, are used. They reached a highest accuracy of 64%. [7] Brazilian lyrics dataset is classified into 6 genres. They used features such as audio signal, lyrics, chords and spectrogram images. Algorithms like K-Nearest Neighbor, Decision Tree, SVM, Random Forest and CNN are used. CNN achieved highest accuracy of 78%. [8] They used a dataset of 200 songs to classify into 20 genres. They compared models by training full song and part of the song. Finally, they observed that using the full song to train the model achieved the best accuracy of 55%. [9] This paper classifies song lyrics based on 6 emotions. They used algorithms like BERT, Naïve Bayes, Random Forest, and a Most Frequent Sense (MFS) model. Embedding techniques of Bag of Words and transformed feature vectors are used. BERT model and Naïve bayes models performed best on different emotion recognitions. [10] This paper classifies the lyrics into 11 genres. LSTM model gave 44.81% validation accuracy and TF-IDF model gave 55.15% validation accuracy. TF-IDF data was applied to logistic regression and testing accuracy of 53.98% is noted.

***Table 2.1:*** *Brief Literature Survey*

|  |  |  |  |
| --- | --- | --- | --- |
| **S.no** | **Authors** | **Title of the Paper** | **Description of model** |
| [1] | Akalp, Hasan, et al. | Language representation models for music genre classification using lyrics | * Models used: BERT, DistilBERT, BiLSTM with both single-label and multi-labels for genre classification |
| [2] | A. Kumar, A. Rajpal and D. Rathore, | Genre Classification using Feature Extraction and Deep Learning Techniques | * Embedding: Bag of Words, TF-IDF * Classifiers: SVM, Naïve Bayes, and XGBoost |
| [3] | Zhuang, Y., Chen, Y., & Zheng, J. | Music genre classification with transformer classifier | * Audio clips of 1000 songs in 10 genres. * On split audio clips, applied transformer classifier and sum the respective probability methods. |
| [4] | Araújo Lima, Raul de, et al. | Brazilian lyrics-based music genre classification using a BLSTM network | * Embedding: Word2vec, Wang2Vec, FastText, Glove * Classifiers: SVM Random Forest and BiLSTM |
| [5] | Kumar, Akshi, et al. | Genre Classification using Word Embeddings and Deep Learning | * Embedding: Word2Vec and Word2Vec with TFIDF * Classifiers: SVM, XGBoost, Deep Neural Networks |
| [6] | Ueno, C. L. R. S., & Silva, D. F. | On Combining Diverse Models for Lyrics-Based Music Genre Classification | * Embedding: Bag of Words * Classifiers: SVM, XGBoost Naïve Bayes |
| [7] | Pereira, Rodolfo M., et al. | Representation Learning vs. Handcrafted Features for Music Genre Classification | * Classifiers: SVM, KNN, Decision Tree, CNN etc. * CNN model with Spectograms performed well with 78% accuracy. |
| [8] | Atsız, E., Albey, E., & Kayış, E. | Effective Training Methods for Automatic Musical Genre Classification | * Acoustic analysis to get acoustic features. * Compared models with full song and half song training. |
| [9] | Edmonds, D., & Sedoc, J. | Multi-Emotion Classification for Song Lyrics | * Embedding: Bag of Words, transformed feature vectors. * Classifiers: BERT, Naïve Bayes, Random Forest and Most Frequent Sense (MSF) |
| [10] | Mubeen, M. | Using Deep Learning to Predict Music Genre from Song Lyrics | * Models used: LSTM, TF-IDF and Logistic Regression |

1. **DATASET INSIGHTS**

The dataset we use is from Kaggle machine learning repository. This dataset was originally obtained from scraping the Brazilian website called Vagalume; it contains two csv files of artists-data and lyrics-data. The combined information from these files have many redundant columns, which are not useful for our project. Such as artist name, song name, song link, artist link, language etc. There were over 79 Genres and many languages in the original dataset. But, we wanted to focus on English songs in only 3 genres: Rock, Pop and Hip-hop.

**4. DATA PRE-PROCESSING**

**4.1 Feature selection**

We extract the required features needed for out project from the original dataset. For this, we remove all other languages except English. Furthermore, we plan to classify only three of the most liked genres: Rock, Pop and Hip-hop. So, all the redundant columns are filtered to leave only two columns of Genre and Lyrics.

**4.2 Stop words removal**

From the dataset, the words which doesn’t add any meaning to the sentences are removed. In English, we come across such type of words as articles, prepositions, pronouns, etc. Removing these stop words helps the model to concentrate on the words that hold more value.

**4.3 Lower case conversion**

The Lyrics column of dataset is converted into lower case for information processing and parsing. Additionally, this can help the model to understand a word’s syntactic role in later stages.

**4.4 Padding the sentences**

All the sentences in Lyrics column are not of equal length. So, we first find out which Lyric has the most number of sentences, and pad all the remaining sentences with zeroes to make them equal.

**4.5 Tokenization**

We use tokenization to split the paragraphs and sentences into words. This process allows the model to understand the whole meaning behind the sentences by reading its individual parts.

**5. METHODOLOGY**

**5.1 USE**

Universal Sentence Encoder (USE) model encodes the text data into high dimensional vectors which are called embeddings. These are the numerical representation of text data. We use the pre-trained USE model which is publicly available in tensorflow-hub. This model takes input according to variable-length English text given, and output is a 512-dimension vector. It comes with two variations one is a transformer encoder model which has a higher accuracy and is computationally more expensive. While the other model is trained with a Deep Averaging Network which is computationally less expensive and has a little lesser accuracy. So, for our project we opt DAN version of USE.

**5.2 BERT**

Bidirectional Encoder Representation from Transformers (BERT) is a language representation model developed by Google AI Language. This model is developed such that it can be bi-directionally trained. It is observed that such a model has a very good knowledge of the context and flow than the single direction language models. BERT makes use of transformer model which learns the contextual relationship between words in a text. In BERT only encoder mechanism is used. Here, the entire sequence of words is taken at once; it studies the context of each word based on its surrounding words. Researchers included a technique called MaskedLM (MLM) in BERT. Due to this, the model converges slower than other directional models. And bidirectional training outperforms the traditional single directional language models.

**5.3 LSTM**

LSTM contains input layer, hidden layer and output layer. In the hidden layer, it has recurrent units just like RNN. But unlike the standard feed forward neural networks, LSTM has a feedback connection. Hence, it can easily process single data points and also entire sequence of data. LSTM architecture aims to provide short term memory for RNN which can last for thousands of steps. The recurrent unit in LSTM is much more complex than the one in RNN which improves learning but more computational resources. This module has a cell state and three gates (input, output and forget gates) which provides the cell states to learn, unlearn and retain information from every unit. This helps the information to flow through the units without being altered and allows only a few linear interactions. The input gate controls flow of information; forget gate removes information from previous cell state using a sigmoid function. The output gate decides which information should be sent to the next hidden stage.

**6. RESULT ANALYSIS**

To develop a model with best accuracy, we have used two different language representation models Universal Sentence Encoder (USE) and Bidirectional Encoder Representation from Transformers (BERT). We chose Long Short Term Memory (LSTM) model for classification purpose. We trained the Language representation models USE and BERT over five folds to determine the best accuracy.

* As shown in Table 2 we have tabulated the best accuracy in each fold for both USE and BERT models.
* We can observe that the best accuracy with USE-LSTM model is 83.42% in Fold 2 and for BERT-LSTM model the best accuracy id 76.55% in second Fold.

***Table 2:*** *Table of Accuracies*

|  |  |  |
| --- | --- | --- |
|  | **USE** | **BERT** |
| **Fold 1** | 79.23% | 64.49% |
| **Fold 2** | 83.42% | 76.55% |
| **Fold 3** | 79.4% | 74.54% |
| **Fold 4** | 80.23% | 75.38% |
| **Fold 5** | 79.23% | 64.15% |

* As displayed in Table 3 we have also analyzed the total parameters for both models. We observed that USE has more parameters than BERT.
* Here we can conclude that even though BERT has less parameters than USE the complexity of USE enables it to learn efficiently.

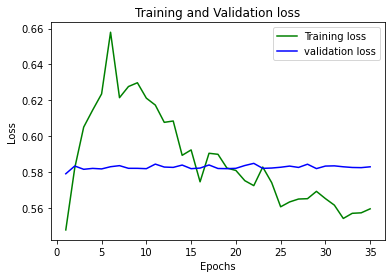
***Table 3:*** *Table of Parameters*

|  |  |  |
| --- | --- | --- |
|  | **USE** | **BERT** |
| **Total Parameters** | 142,851,170 | 36,406,370 |
| **Trainable Parameters** | 142,851,170 | 36,406,370 |

* In Fig 6.1 and Fig 6.2 we can view the training and testing loss of both USE and BERT models.



***Fig 6.1:*** *Training and Validation loss of USE*

**

***Fig 6.2:*** *Training and Validation loss of BERT*

**7. CONCLUSION**

In Conclusion, in this paper compared various language models to determine the best model. Here we have developed a song genre classification model suing lyrics with Language Representation models and Recurrent Neural Networks with a highest accuracy of 83.42%. This model will provide an automatic genre classification to manage the songs according to its genre in music platforms.

**8. REFERENCES**

[1] Akalp, Hasan, et al. "Language representation models for music genre classification using lyrics." 2021 International Symposium on Electrical, Electronics and Information Engineering. 2021.

[2] A. Kumar, A. Rajpal and D. Rathore, "Genre Classification using Feature Extraction and Deep Learning Techniques," 2018 10th International Conference on Knowledge and Systems Engineering (KSE), 2018, pp. 175-180, doi: 10.1109/KSE.2018.8573325.

[3] Zhuang, Y., Chen, Y., & Zheng, J. (2020, June). Music genre classification with transformer classifier. In Proceedings of the 2020 4th International Conference on Digital Signal Processing (pp. 155-159).

[4] Araújo Lima, Raul de, et al. "Brazilian lyrics-based music genre classification using a BLSTM network." International Conference on Artificial Intelligence and Soft Computing. Springer, Cham, 2020.

[5] Kumar, Akshi, Arjun Rajpal, and Dushyant Rathore. "Genre classification using word embeddings and deep learning." In *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 2142-2146. IEEE, 2018.

[6] Ueno, C. L. R. S., & Silva, D. F. (2019, October). On Combining Diverse Models for Lyrics-Based Music Genre Classification. In *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)* (pp. 138-143). IEEE.

[7] Pereira, Rodolfo M., et al. "Representation learning vs. Handcrafted features for music genre classification." *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019.

[8] Atsız, E., Albey, E., & Kayış, E. (2019). Effective training methods for automatic musical genre classification. In *DATA 2019-Proceedings of the 8th International Conference on Data Science, Technology and Applications*. SciTePress.

[9] Edmonds, D., & Sedoc, J. (2021, April). Multi-emotion classification for song lyrics. In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp. 221-235).

[10] Mubeen, M. Using Deep Learning to Predict Music Genre from Song Lyrics.