ISMTS-117 – Text Analytics and Natural Language Processing

Final Project
Music Genre Classification of Lyrics using LSTM
Narayana Bandhi

1 RESEARCH QUESTION

Automatic classification of music is an important and well researched task in music information retrieval (MIR). Genre classification by lyrics presents itself as a natural language processing (NLP) problem. In NLP the aim is to assign meaning and labels to text; here this equates to a genre classification of the lyrical text.

SVMs, k-NN, and NB have been heavily used in previous lyrical classification research. These models produced classification accuracies of 69% amongst 5 genres and 43% amongst 10 genres.

Can an LSTM bi-directional model using GloVe embeddings improve the accuracy of genre classification?

2 UNDERSTANDING THE DATA

2.1 DATA SOURCE

The data comes from Kaggle that has a dataset of artists and lyrics. There are about 209K lyrics across 6 genres and multiple languages. For the purpose of the project I only use English language lyrics around 109K across 3 genres.
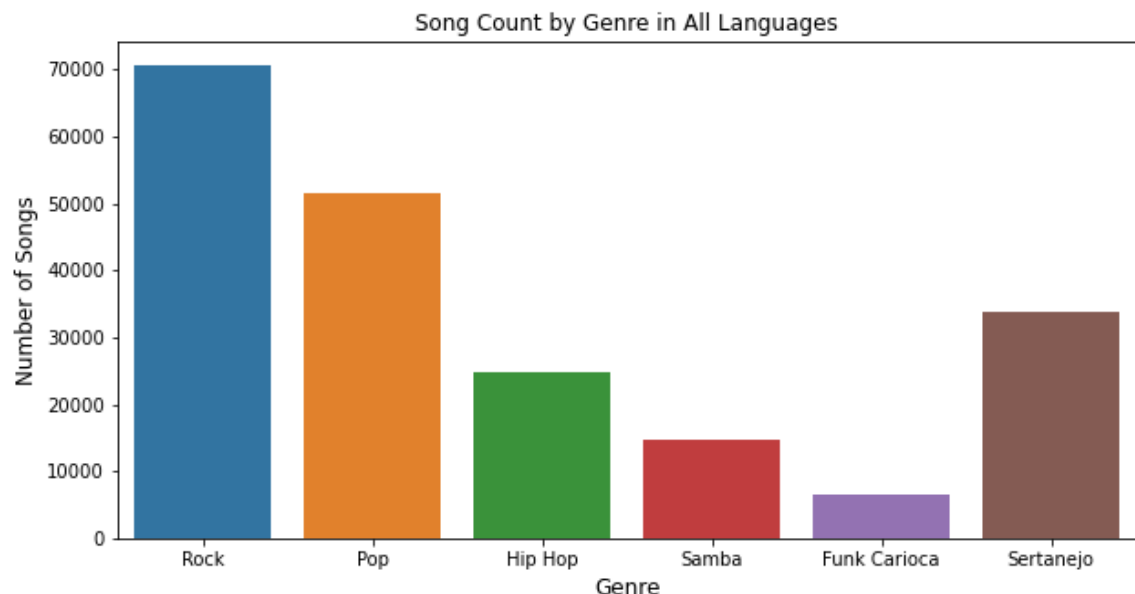
2.1.1 Exploring the Genres

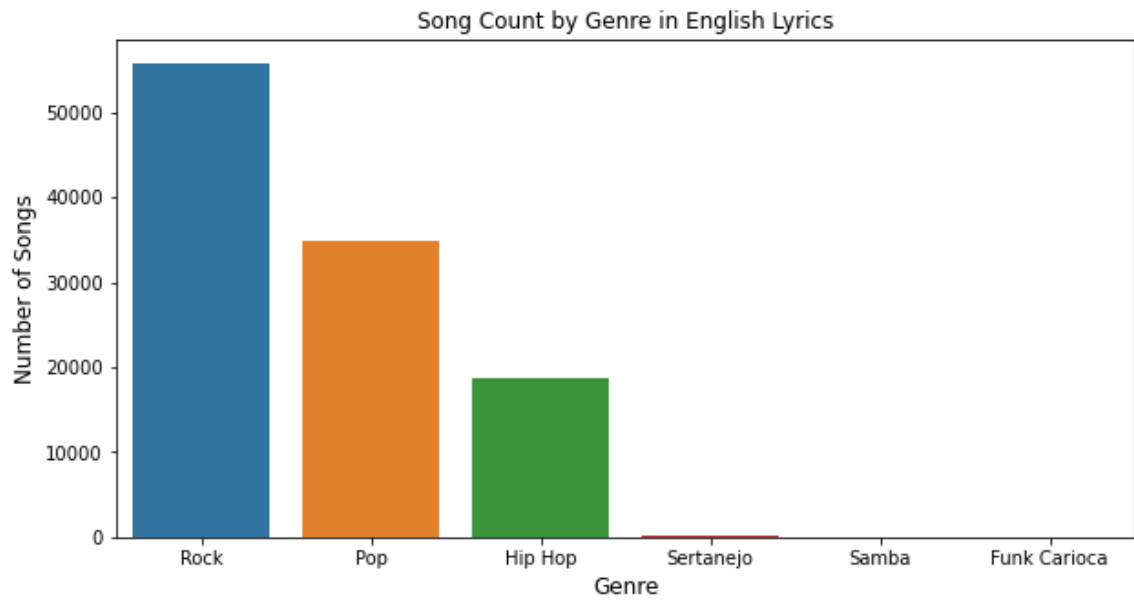Fig. 1 – Genres across all Language Lyrics

Fig. 2 – Genres across all English Language Lyrics
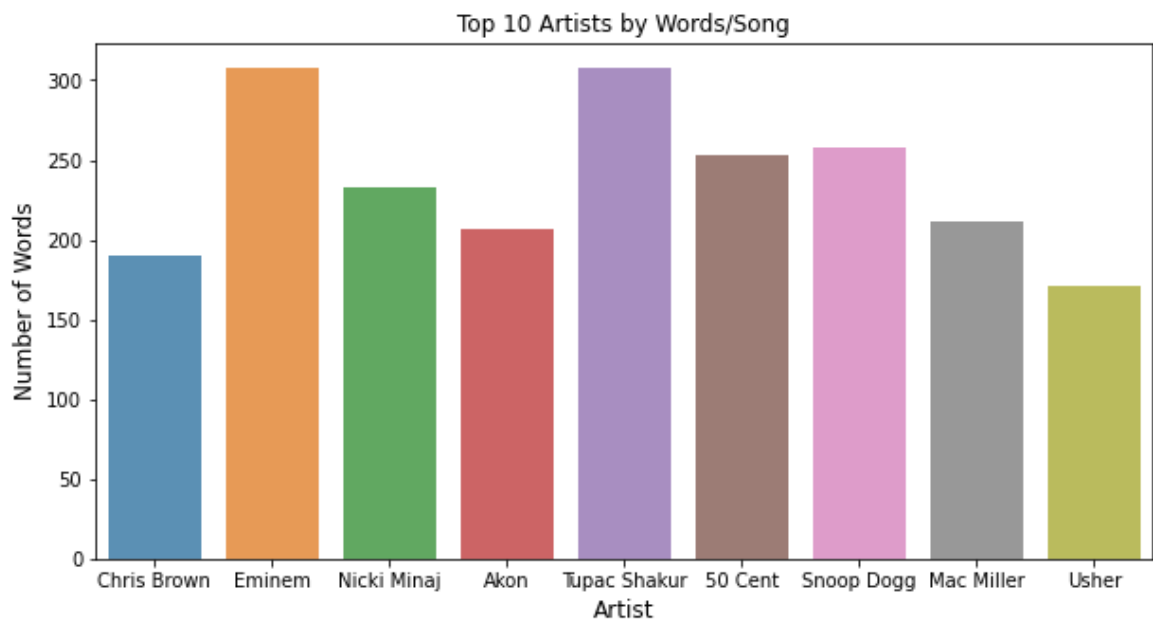
2.1.2 Exploring the Word Counts

Fig. 3 – Top 10 Artists by Words per Song

Fig. 4 – Word Vector – Rock Lyrics

Fig. 5 – Word Vector – Pop Lyrics

Fig. 6 – Word Vector – Hip Hop Lyrics

2.1.3 Topic Models

Used LDA to identify various topics and word associated with the topics.

3 MODEL

The motivation to use Long Short Term Memory model stems from the fact that lyrics are inherently sequential in nature, and the similarity between two lyrics must in at least some way be determined by the similarities between their sequences over time.

An important idea in NLP is the use of dense vectors to represent words. For the purpose of this project I will be using GloVe embeddings.

To address overfitting since the corpus is 5000 lyrics I will incorporate dropout rates.

LSTM (Long Short-Term Memory) are very good for analyzing sequences of values and predicting the next one. For example, LSTM could be a good choice if you want to predict the very next point of a given time series.

Talking about sentences in texts, the pharse (sentences) are basically sequences of words. So, it is natural to assume LStM could be useful to generate the next word of a given sentence.

In summary, the objective of our LSTM neural network will be to predict the genre given the lyrics of a song.

4 EXPERIMENT

As an initial step calculated accuracy of predicting the genre using SVM and Count Vectors. The results are shown in Table 1.

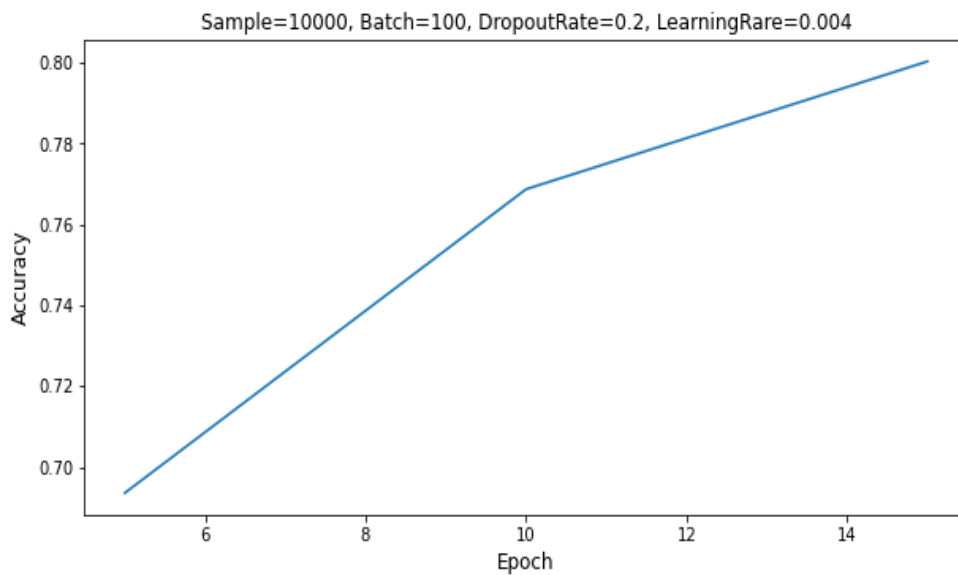| Sample Size	Accuracy
1000	0.63
5000	0.683
10000	0.693
109000	0.114

Table 1 – Accuracy using Count Vectors and SVM

We noticed that the the accuracy increased with the sample size but then it decreased when tested across the full corpus of lyrics of 109K songs.

We ran the data through LSTM model and calcutated the accuracy for various samples size for epochs of 5, 10 and 15. The results are shown in Table 2 and Fig 7.

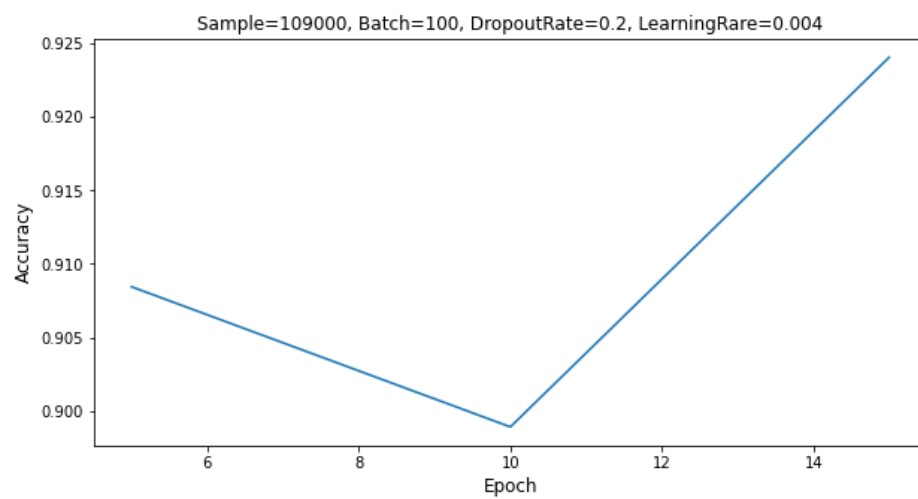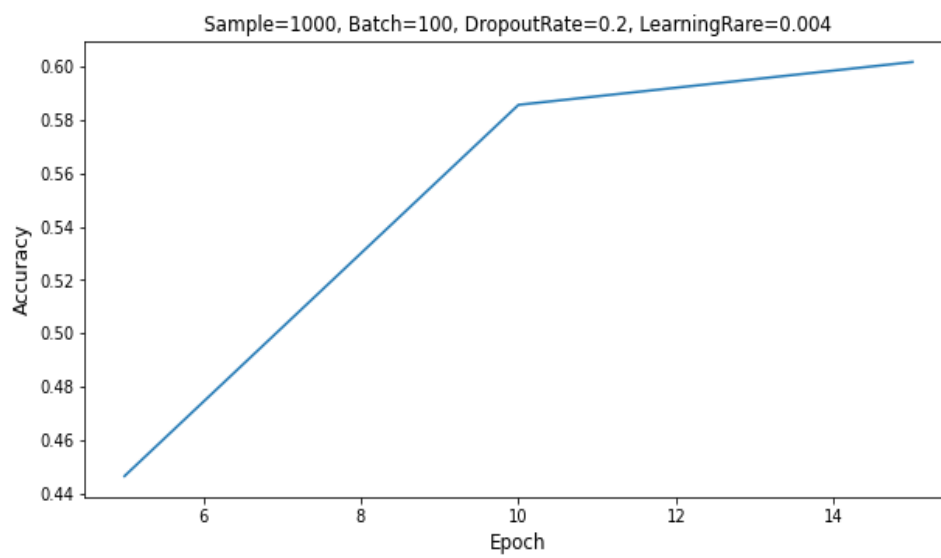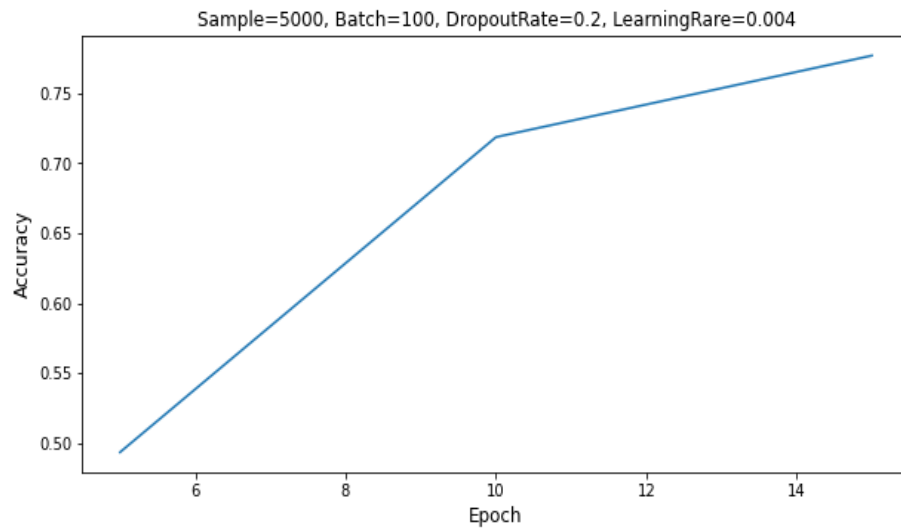| Sample Size	Epoch 5	Epoch 10	Epoch 15
1000	0.4464	0.5857	0.6017
5000	0.4935	0.7185	0.7767
10000	0.6935	0.7687	0.8003
109000	0.9084	0.8989	0.9240

Table 2 – Accuracy using LSTM and GloVe Vectors

Fig 7 – Accuracy across various epochs for different sample sizes

5 DISCUSSION

The LSTM model had a better accuracy rate in predicting the genre of lyrics even with a small corpus size of 5000 and it siginificantly increase with a larger corpus size of 10000 lyrics.