Ryan Rey M. Daga

Project Genre X: Music Genre Classification using N-gram Classes of Lyrics

I. Introduction

Music genre classification is a computing problem that predicts the genre of a music of a given audio data. Additionally, lyrics provide additional information for music genre classification. In this research, music lyrics were utilized in this study to predict the genre of an input text. Text processing techniques and Neural network - based models were employed to build a music genre classifier using n-gram classes of the lyrics as features. The way n-grams were utilized as features to define the characteristic of a lyric will be discussed in the section ahead.

II. Proposed Music Genre Classification using N-gram Classes

The process flow of the proposed music genre classification is shown in Figure 1. The details of each phase will be discussed ahead.

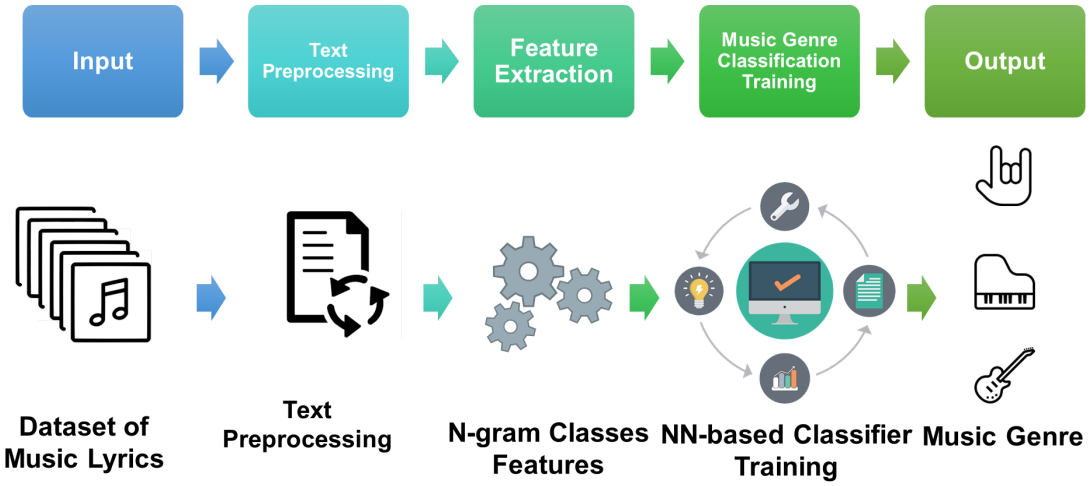


Figure 1. General flow of the proposed music genre classification using lyrics

A. Input

The dataset of music lyrics utilized in this research was collected from Mendeley Data entitled Music Dataset: Lyrics and Metadata from 1950 to 2019. The collection contains 28,372 song lyrics with 7 music genres: (1) pop, (2) jazz, (3) blues, (4) reggae, (5) hip hop, (6) rock, and (7) country.

B. Text Preprocessing

Different text processing techniques were applied to the lyrics of the dataset to ensure that better features were extracted by succeeding processes in the music genre classification. The first preprocessing technique was to remove

punctuations and non-letter characters in the lyrics. Next, stop words were removed to ensure that only significant tokens remain in the dataset. Lastly, lemmatization was applied to each token to extract the base word of each text in the lyrics.

C. N-gram Classes Feature Extraction

Feature extraction was performed to each lyric in the dataset to get a mathematical representation of the song lyrics. For each song lyric, n-gram is extracted and the frequency of occurrence of each n-gram in the entire document is determined.

ng_1	ng_2	ng_3	ng_4	ng_5		ng_l
2	150	3	20	90		35

Figure 2. Example of N-gram its corresponding frequency of a song lyrics

Then, the maximum frequency from the n-gram document model is determined. It will be utilized to determine the base which is computed by determining the (***m***-1)*th* root of the maximum frequency where ***m*** is the predetermined number of n-gram classes. Figure 3 shows a sample computation of the base from the n-gram document model.

ng_1	ng_2	ng_3	ng_4	ng_5		ng_l
2	150	3	20	90		35

max_freq

$m = 3$

$base = \sqrt[2]{150} = 12.25$

Figure 3. Computation of base from n-gram document model

Using the base computed earlier, the n-gram class of each n-gram in the document model is computed by determining the logarithm of its frequency.

ng_1	ng_2	ng_3	ng_4	ng_5		ng_l
2	150	3	20	90		35

$m = 3$ $base = \sqrt[2]{150} = 12.25$

$class\ ng4 \cong \log_{12.25}(20)$

ng_1	ng_2	ng_3	ng_4	ng_5		ng_l
0	2	0	1	2		1

n-gram class document model

Figure 4. N-gram class computation for each n-gram

Lastly, distribution of n-gram classes is determined by counting the number of n-grams that belong in less frequent, intermediate, and high frequency n-grams.

ng_1	ng_2	ng_3	ng_4	ng_5		ng_l
0	2	0	1	2		1

n-gram class document model

class 0	*Class 1*	*Class 2*
0.75	0.1	0.13

Figure 5. Example n-gram class document model

This process is performed to each row of music lyrics in the entire dataset. Then the genre of each music lyric is mapped to a numeric integer. The features extracted are contained in a dataframe containing the n-gram class document model and its corresponding genre.

D. Music Genre Classification Training

The training process of a Neural Network based model usually takes a long period of time. To be able to build rapid prototypes of classifiers with different combinations of n-grams size and n-gram class number, Naive Bayes classifier was used to build initial model classifiers. To train the model, a 70:30 ratio was utilized to split the training data and testing data.

III. Results and Discussion

The proposed music genre classifier was built using Python.

A. Preliminary Results

Initial classifiers were built using Naive Bayes to be able to quickly build and test different combinations of n-gram size and n-gram class size. The accuracy of classification is shown in Table 1 and visualized in Figure 6.

Table 1. Accuracy of Naive Bayes-based Classifier

n-gram size	n-gram classes			
	3-class	4-class	5-class	6-class
bigram	0.19843	0.17937	0.19543	0.19655
trigram	0.20141	0.20928	0.19395	0.21450
4-gram	0.20290	0.20032	0.20066	0.19805
5-gram	0.19955	0.21340	0.19542	0.20928

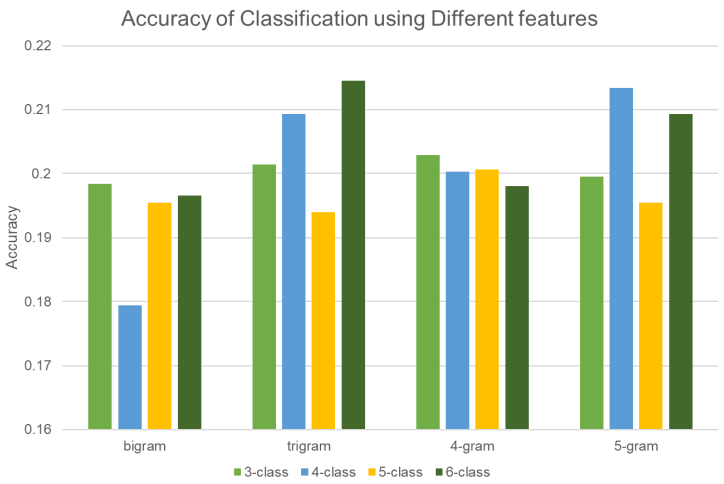
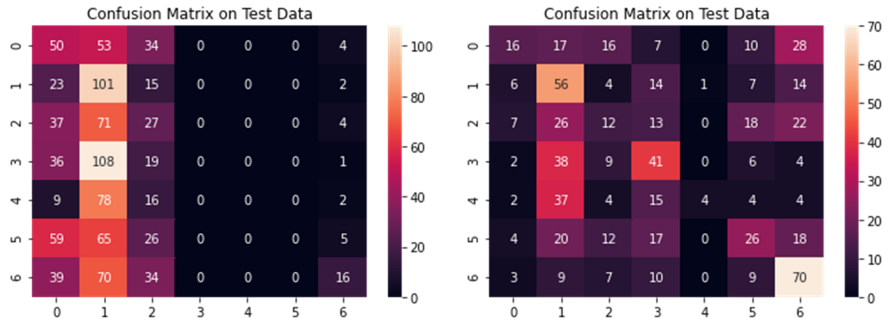


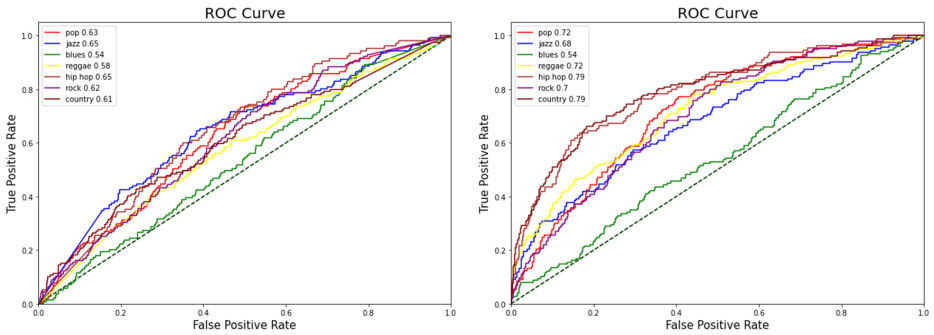
Figure 6. Accuracy of Naive Bayes-based models using different n-gram size and n–gram class size

The performance of the classifier using the proposed feature was compared to TF-IDF as a baseline. Figure 7 shows the confusion matrix in a heatmap form to evaluate the two.



N-gram Classes **TF-IDF**

Figure 7. Heat map of the confusion matrix of proposed feature and TF-IDF



N-gram Classes **TF-IDF**

Figure 8. ROC Curve of the classifier using proposed feature and TF-IDF

ROC Curves was also generated to see the performance of the classifier for each individual music genre as shown in Figure 8.

B. Neural Network-based Classifier

After building prototypes of the classifier using different combinations of n-gram size and n-gram class size, a Neural Network-based classifier was built to evaluate its performance. The n-gram size utilized was 4 and the n-gram-class size was set to 3, and 5 hidden layers for the perceptron. The NN-based classifier that utilized n-gram class features yielded lower accuracy compared to TF-IDF.

Table 2. Accuracy of classification of the NN-based classifier

Feature	**Accuracy**
n-gram class	21.7%
TF-IDF	24.3%

IV. Conclusion

A neural network-based classifier that utilized n-gram class as features was proposed in this work. N-gram classes as a feature for genre classification using lyrics produced a slightly lower result compared to TF-IDF. However, this feature can be computed independently for each document and a shorter feature size can be carried out. The proposed n-gram class feature can be utilized to supplement lyric features towards music genre classification and may be combined with other song features for a better classification performance.