Lyric-based Classification of Musical Genres

Brian Ha, Kevin Deng, Timothy Yang, Tion Wong
Boston College Morrissey College of Arts and Sciences: CSCI 3349.01 Fall 2019 Natural Language Processing

Background

- Song genre classification has applications in browsing and searching music databases, particularly song recommendation systems
- We hypothesize that...
  - Genres are separable by the language and range of words that they use
  - Genres exhibit notable differences in topics and can be classified
- Goal: determine whether song genres can be predicted through song lyric data

Data

- 380,000+ lyrics from MetroLyrics in the format of Lyrics, Artist, Genre, and Year
- **Data Preprocessing**
  - Omitted songs without lyrics and songs not in English using python library WhatTheLang
  - Removed/replaced problematic lyric fields (e.g null values, too few words, symbols etc.)
  - Removed the 5 genres with the fewest songs
  - Unbalanced (labeled) → Balanced (labeled) dataset
    - ⇨ Took a random sample of size n=13354 for each of the 5 remaining genres
    - ⇨ n = size of the smallest remaining genre

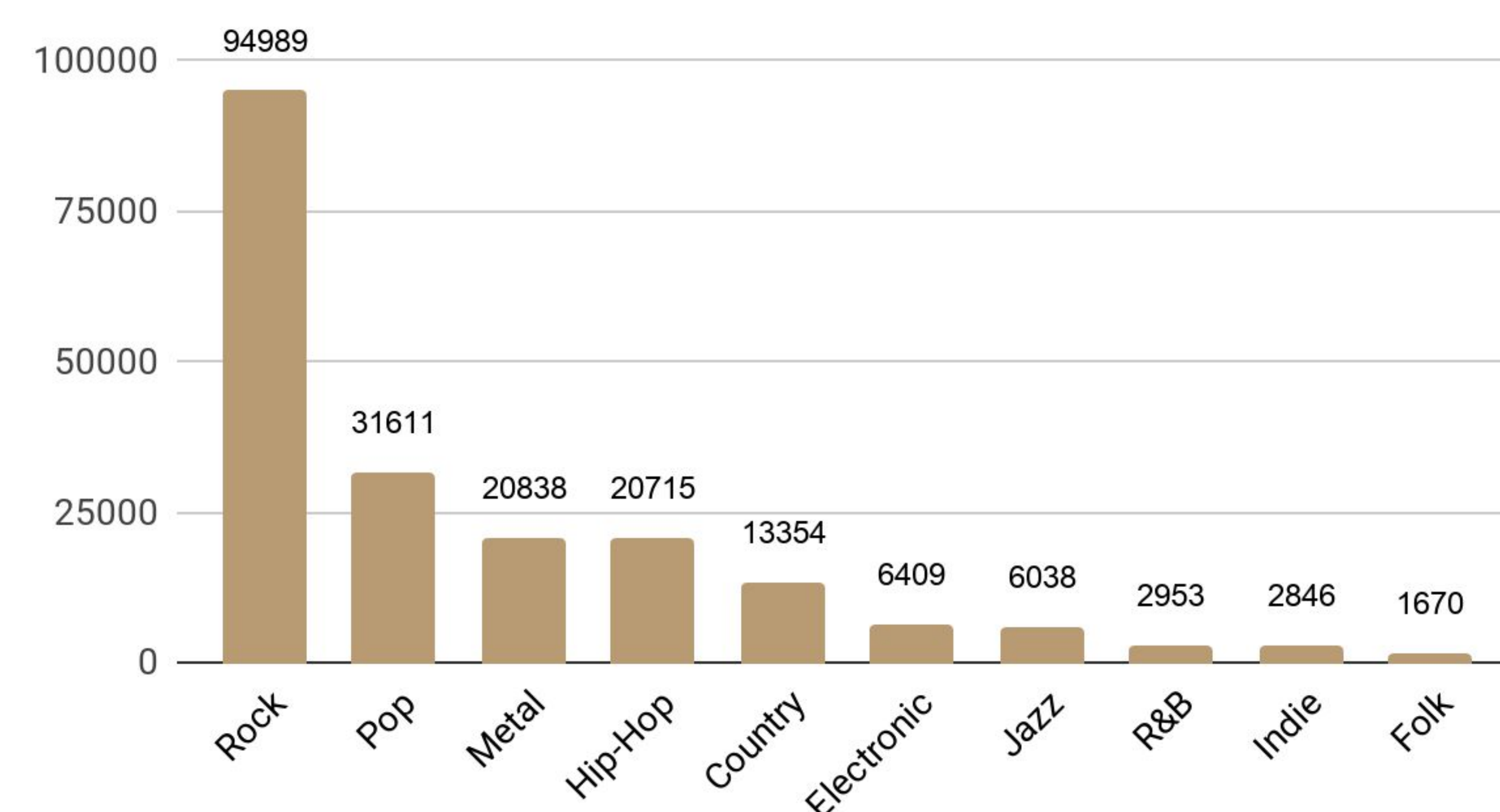| Dataset Name	# Songs
unbalanced-labeled	201423
balanced-labeled	66765

Fig. 1: Number of songs per genre of *unbalanced-labeled* dataset

Methodology

- **Multinomial Classification Techniques**
  - Naïve Bayes (NB)
  - Logistic Regression (LR)
  - k-Nearest Neighbors (k-NN)
  - Support Vector Machines (SVM)
  - Long Short-Term Memory Recurrent Neural Networks (LSTM RNN)
- **Classification Process (NB, LR, k-NN, SVM)**
  - Removed stopwords and lemmatized tokens for classification
  - tf-idf vectorization was performed to vectorize and transform lyric data
    - ⇨ We set the max number of features to be 10,000 (large vector)
  - We split the resampled dataset into 80% training, 20% testing
  - Trained the 4 classifiers on the data, then calculated accuracy.
- **Classification Process (LSTM RNN)**
  - 80% of the training/validation set was used for training, and 20% used for validation (improving the model)
  - Used the fast.ai library to perform Transfer Learning using a language model pre-trained on the "Wikitext-103" dataset
  - Fine-tuned the model to fit the dataset, and used the resulting model to make genre predictions on test data and generate a confusion matrix

Results

Classification Technique	Accuracy
Baseline	0.20000
Naïve Bayes	0.63050
k-NN	0.48506
Logistic Regression	0.64794
SVM	0.63102
LSTM RNN	0.67947

Fig. 2: Accuracies for the five classifiers

- We decided that accuracy was a good measure since our dataset was balanced, and every genre was equally important to classify
- Not too surprising that RNN achieved best result since it uses a pre-trained language model which we fine tuned to our dataset
- By analyzing the resulting confusion matrices, it is clear that all of our models had very similar strengths and weaknesses:
  - Hip-Hop was the category in which our classifiers performed the best in by far
  - Rock and Pop were often misclassified for each other as well as Rock and Country, Metal and Rock
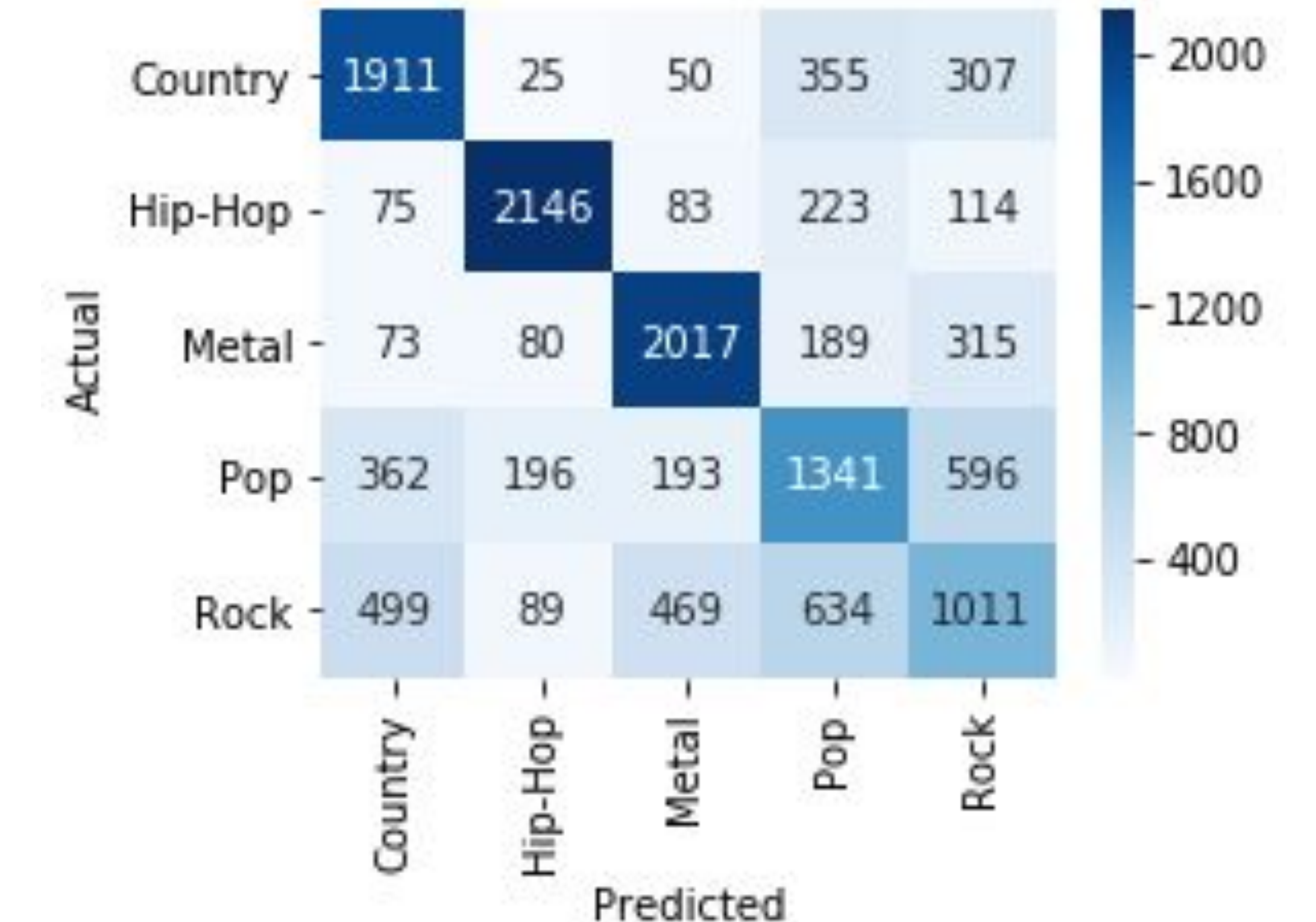
Results (continued)

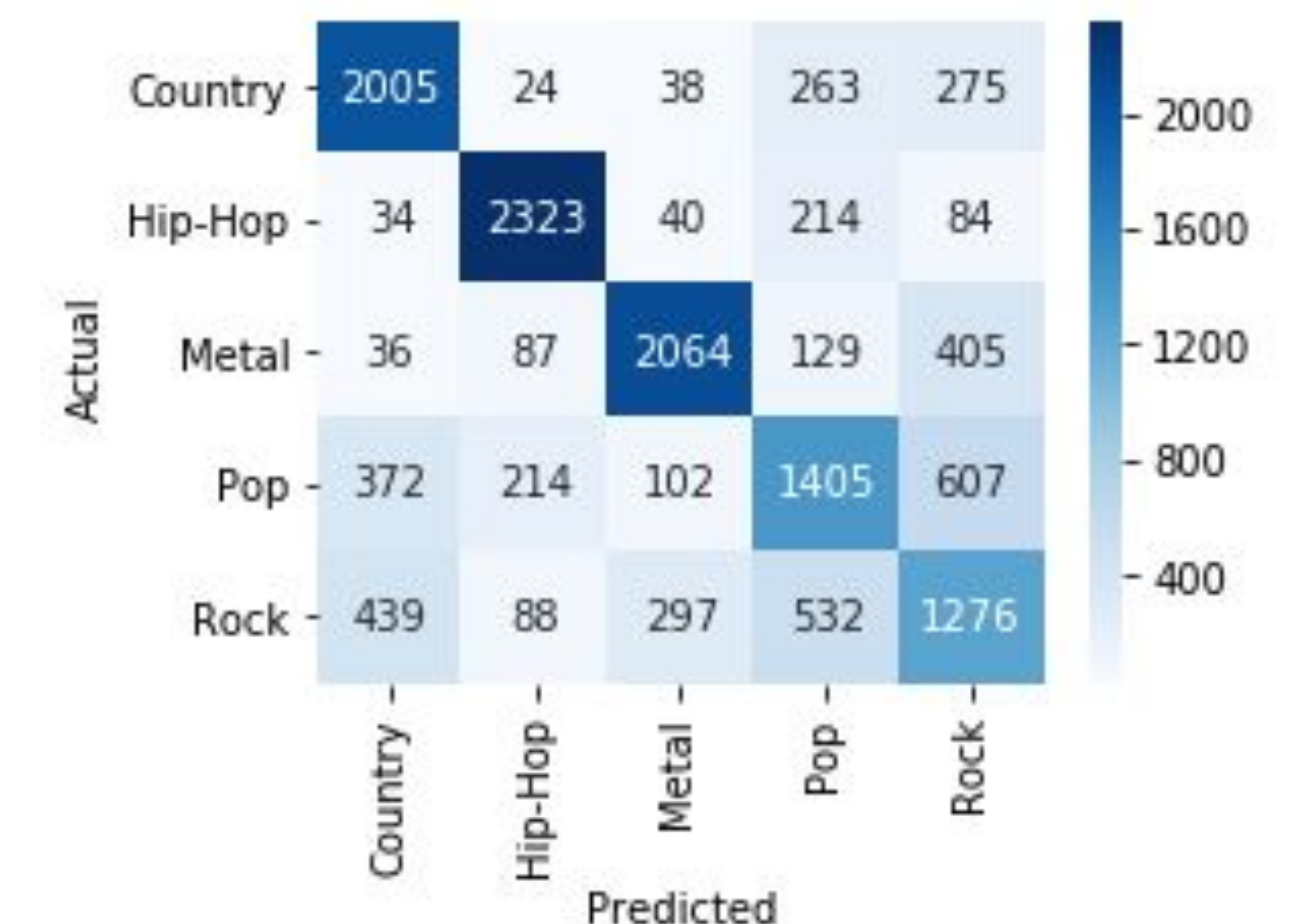Fig. 3: Confusion Matrix for SVM

Fig. 4: Confusion Matrix for RNN

Conclusions and Future Work

- Results show that it is possible to classify musical genres based on song lyrics with limited performance
- Potential factors
  - Song lyrics are not a good predictor of genre, as many genres are intertwined and similar in content
  - Flaws in dataset - lyric formats are not consistent and language detector had performance issues
- Future work
  - Train on unbalanced dataset to mirror genre-wide volume differences
  - Experiment with different test/train splits/preprocessing techniques
  - Perform more dataset analysis and find which words or topics correspond to which genres with word clouds