HKUST MSBD 5002: Generating Rap Lyrics with Deep Learning Approaches
Group Proposal - Implementation Project

Ko Wing Fung : 20279217
Koo Tin Lok : 20344775
Shiu Stephen : 20041373
Muhammad Adam Usmani: 20729858
Qiwei Li : 20720230
Ng Pui Yan : 20670370

1 Introduction

Rapping emerged from a hobby of African American youth in the 1970's, it quickly evolved into a mainstream music genre with several artists frequenting Billboard top rankings.

Writing high quality rap lyrics requires creativity, to be able to construct interesting-story-telling, and-also lyrical skills, which is the foundation behind forming rhythmic complexity. In our paper, we will be proposing several deep learning methods to tackle rap lyric-generation. Our goal is to capture the patterns and styles of multiple lyricists by being able to generate new lyrics.

It is our decision to study the computational creation of rap lyrics since this is a genre of music which distinguishes itself from other genres. Rap lyrics possess a formal structure which is made up of very carefully crafted sound and rhyme. Our focus is to analyze this formal structure and develop models that are-able-to generate the same artistic style that is present in multiple rap songs.

2 Dataset

For the data, we would only need the text-base lyrics. We expect the dataset will just be a collection of song lyrics lines.

There is public dataset [1] on Kaggle already, but the size is small, that dataset contains only 500 songs only. As a result, we choose to gather our own data by scrapping data from "OHHLA website [2]" The website archives English rap lyrics from 3604 rappers starting from 1992. The total number of rap song in this website is not known to us, but we would figure out the size of dataset after the scrapping process. The website organized hip hop song lyrics in a structured hierarchy – rapper/album/song. Each lyrics website is just a plain text file of the lyrics. It should take us only a little effort to extract all the lyrics text from the website.

We would need to clean the lyrics after scrapping, because the lyrics file contains some unnecessary information tags such as the artists name, album name, song name, etc. Meanwhile, because the lyrics are uploaded and maintained by different people, there are different format of the songs, an example would be having extract tags on the lyrics such as "[Verse]", "[Chorus]" that are not useful to us. Anyway, the scrapping and cleaning processing should take only a very short amount of time.

3 Methodology

The main challenge for rap lyrics generation is expected to be the rhyming and semantic meaning. For rhyming, there are various rhyme types, such as "perfect rhyme", "alliteration" and "consonance" while "assonance" is the most common rhyme type nowadays. In particular, perfect rhyme highlights the words share exactly same end sound, while "assonance" highlights only the vowels sounds are shared. In this sense, rap lyrics generation

[1] Rap Songs (data from Genius) - Kaggle https://www.kaggle.com/sebastianrs95/rap-songs-data-from-genius

[2] The Original Hip-Hop (Rap) Lyrics Archive - OHHLA.com - Hip-Hop Since 1992 http://ohhla.com/all.html

differs from text generation. The latter one focuses on correlation between words (e.g. the N-gram model) and the semantic meaning of overall sentence, while the former one emphasizes the pronunciation that is hard and exhaustive to quantify and the semantic meaning. Apart from the rhyme, the sentence length in each sentence, the popularity of the selected words, the semantic meaning between lines are possibly key factors to assess "The produced lyrics make it a good rap or not?". Although it may also relate to the subjective elements like the rapper's style, they all indicate that our knowledge of all possible "good" features can never be comprehensive. Using a computational approach to assess a rap song appears to be a convincing method. Therefore, we target to build a discriminator aiming for "Given a sentence of lyrics to distinguish the next line of lyrics is a suitable lyric or not". In the way of building a discriminator is expected to involve text preprocessing, language modeling, and other possible machine learning models like neural network, hence the network architecture tuning. Moreover, due to the large amount of dataset, the final configuration must take the limited processing power and time constraint as consideration to strike for the optimal result.

Having the standard of a good rap song can only let you know more about the mystery. We target to generate a sound lyric. We expect word-by-word rap lyric generation can be challenging because 1. Quite a number of words in rap lyric is very rare and hard to build a language model 2. Having a rhyme word without deforming the original meaning of whole sentence requires huge dictionary. Therefore, we will define the aim as "next-line prediction" first because the existing written rap lyrics is exactly a suitable repository. We purpose to build a generator. In the process, we need to extract the features of the sentence. The expected ones include the length of lyrics, rhyme density, similarities between sentences, semantic meaning. Various tools like BOW, Jaccard similarity, LSA, NN5 are potential candidates for feature extraction. A machine learning model (e.g. RankSVM) will be built to select a good line from pools of sentence. Only the sentence that passes the model as well as passing the test of "rhyme_ok" will become the final candidate. It is possible to make use of various tools like "epitran" or "pyrhymes" for rhyme detection. Some more feature extraction includes the identification of longest matching vowel sequence. They become an option which is likely to enhance the quality of generation depending on the time.

4 Preprocessing

Before building our models, preprocessing the lyrics is extremely important. The objective is to transform the lyrics into a more digestible form for the models and make the result from the lyrics generator to be much more accurate. Here we planned there are at least four steps to do. Firstly, we would do lower casing for all lyrics. It is to make sure the original words with different casing such as "Love" and "love" are representing the same word. The second step is to convert the digits into text such as converting "12" to "twelve". The motivation here is similar with the one in lyrics lower casing. With the third step, we clean the dataset and remove the meaningless special characters in it so we will preprocess with the special characters that we only preserve some of them such as spaces, semi-colons, commas etc. The last step we would like to remove non-english lyrics so we could focus on generating the English lyrics only.

5 Evaluation

For most of the Natural Language Generation (NLG) task, evaluation is challenging because the task is open-ended. It is difficult to assess the performance of the computational approaches as there is not a 'ground truth' of rap lyrics. Besides, it is subjective when it comes to evaluating lyrics. While for typical rap lyrics, rhyming is usually one of the features inside a rap song. As a result, we will try to quantify the technical quality of lyrics from a rhyming perspective. Rhyme density measure is introduced to evaluate the lyrics as follows:

First, the generated lyrics will be pre-processed by computing the phonetic transcription of the lyrics and removing all but vowel phonemes. Then we scan the lyrics word by word. For each word, we find the longest matching vowel sequence that ends with one of the 15 previous words. Starting with the last vowels of two words, if they have the same vowels, we proceed to the second to last vowels, third to last, and so on. We proceed, ignoring word boundaries, until the first non-matching vowels have been encountered. Finally, we compute the rhyme density by averaging the lengths of the longest matching vowel sequences of all words.

Apart from the rhyme density measure, we might also perform evaluation via lexical diversity. It is one of the commonly used metrics in NLG task, to measure how many different words are used in the lyrics.

While human evaluation is regarded as the gold standard in NLG task evaluation, we will also conduct an evaluation to score the quality of the generated lyrics based on their fluency, coherence and correctness, to try to compare the evaluation from the human and automatic metrics perspective.

6 Expected Outcome

The outcome for this project is using the methods above to auto generate a set of lyrics. We expect the final app can work as the following: When user provide a set of sentences and the program can generate this based on this initial input. If time permit, we can improve this application so that it can generate lyrics in different styles (ex: Country music, pop music, and Hip-pop music). Since the text (lyrics) generation can be replaced to different styles once we have the initial model set up. In order to share our final product to others and can be used in real world, we will build a website to let user to utilize our model.

References

1. DopeLearning: A Computational Approach to Rap Lyrics Generation https://www.kdd.org/kdd2016/papers/files/adf0399-malmiA.pdf
2. GhostWriter: Using an LSTM for Automatic Rap Lyric Generation https://www.aclweb.org/anthology/D15-1221.pdf
3. Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling https://static.googleusercontent.com/media/research.google.com/zh-TW//pubs/archive/43905.pdf

DECLARATION STATEMENT

MSBD5002 Data Mining and Knowledge Discovery

Group 5

We declare that this project is being done solely within this course and no other scopes.

Candidate Agreements:

Student 1: Ko Wing Fung: 20279217

Student 2: Muhammad Adam Usmani: 20729858

Student 3: Koo Tin Lok: 20344775

- Your FYP supervisors: Prof. Bertram SHI (ECE, HKUST) & Prof. Xiaojuan MA (CSE, HKUST)
- FYP Topic: Automatic 3D Head Modeling

The project is about image processing, not related to this project at all

- This project is done solely within the course but-not other scopes

Student 4: Qiwei Li: 20720230

Student 5: Shiu Stephen: 20041373

Student 6: Ng Pui Yan: 20670370