A MODEL DESIGNED FOR AUTOMATIC GENERATED RAP LYRICS IN GIVEN GENDER AND STYLE 2021 LBD

Gu Lingan
Nanyang Technology University

ABSTRACT

Automatic lyrics generation with deep learning has been a popular research area. The methods have evolved from rule-based approaches into AI-driven systems. Existing approaches, however, are mainly focused on alignment of melody and rhythm, or ignore the diversity of different rap styles. Besides, current rap lyrics generators are mostly preconditioned to generate lyrics from male rapper’s perspective, the female voice and female rappers are severely neglected. This paper proposes the lyrics generation model using Generative Adversarial Networks, which can generate rap lyrics based on the input text style and a beat option. Furthermore, the generator allows additional input condition: the rapper’s gender.

1. INTRODUCTION

Rap music, from MC’s improvise in the 1970s club to today’s Billboard hot 100 hits, has emerged from underground voice into mainstream. The spectrum of rap music has contained various schools of rap styles, but the core of this genre remains in its lyrics. Rap is seen as an analogy of poetry in many researchers’ eyes, and similar approaches like test generation and style transfer models have been applied to generate rap lyrics.

Rap has various and flexible rhyme scheme depending on the rap style. However, previous works of rap lyrics generation mainly focus on exploring the rap song’s semantic structure and the interference between rhymes [1]. The models generate rap lyrics by crowd-sourcing fitted words to fixed rhyme schemes. The stylistic feature is often discarded for simplifying the algorithm. Furthermore, many researcher ignores the distinguished features between male and female rappers [2]. It could be said that most rap lyrics generator are presumed to generate only male rapper’s lyrics.

In this paper, we use Sequence Generative Adversarial Networks (SeqGAN) [3] to model lyrics in different rap styles, and eventually use the algorithms to generate samples of lyrics with given beat.

2. DATASET

This work requires two datasets: the lyrics and the beats. The rap beats was originally created by myself containing 12 different beats in 3 styles: Pop Rap Jazz Rap, Trap. We tag the style based on the top rank album chart on Rate Your Music [1]. We then scrap the gender of each artist from Musicbrainz [2], and the lyrics of each song from Genius API [3], and bests from Tunebat [4].

2.1 Lyrics

The lyrics dataset requires information of beat, rhyme, style and gender. However, currently available rap datasets contain only information about lyrics and creditors. Therefore, we designed a pipeline for data mining. The pipeline contain 4 steps: data crawling, beat detection, style selection, gender tagging as Figure 1 shows.

2.2 Beat

The beats are annotated with tempo and style. Based on the lyrics dataset I created the beats according to the BPM that are most frequently used in each style, and imitate the original rap song beat by applying the similar synthesizers and sound effects. Each beat will be embedded in the targeted styles of generated lyrics.

© F. Author, S. Author, and T. Author. Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).


1 https://rateyourmusic.com/
2 https://musicbrainz.org/
3 https://docs.genius.com/
4 https://tunebat.com/
3. APPROACH

The methods used to generate rap lyrics include 4 models: 1 baseline models and 3 conditioned models.

3.1 Bi-gram Model

To generate rap lyrics that are understandable, as well as aligned to the beat, we apply bi-gram model in the process. We compute the probability of a word given the previous words and beat using only the conditional probability if one preceding word. And the evaluation is estimated through MLE.

3.2 LSTM Model Training

The first stage of SeqGAN training is to translate the rap lyrics into integers. During training, the data is converted into parameters. In this case particularly, we will use condition SeqGANs. And the evaluation of the model is based on the lyric generation task’s MLE performance.

3.3 Adversarial Training

In the Adversarial training stage, the generator and discriminator are trained alternately. When training the discriminator, positive examples are from the given dataset, whereas negative examples are generated from our generator. To keep the balance, the number of negative examples we generate for each last step is the same as the positive examples.

3.4 Conditioned Model

To include information on gender, rhymes, and styles, three embedding steps are added separately to combine with lyrics generation. The given style and beat are taken as inputs and lines of lyrics are generated as the output. Figure 2 shows how the lyrics are trained.

4. EXPERIMENT

We divide our datasets into three parts: 800k of the dataset is used for training, 200k is used for validation and 200k is reserved for testing.

4.1 Experiment 1.

We hyper parameterized the embedded length and hidden dimension of the RNN corresponding to the LSTM generator. We tried various values for embedded length (64, 128, 192, 256) and hidden dimension (50, 100, 150, 200) and found that the largest model (embedded length=256, hidden dimension=200) achieves the minimal validation loss. After the best hyper parameters are chosen for each dataset, each model is trained using these hyper parameters and the various evaluation metrics are computed, which we will talk about in the next subsection.

5. EVALUATION

The generated rap lyrics are evaluated with the BLEU score for the similarity between the generated lyrics and the original lyrics. The alignment ratio test is to examine the lyrics are aligned with the beat and rhyme. And the style will be hand evaluated since the suitable evaluation metrics for rap genre are still lack of exploration.

6. RESULT

For LSTM-RNN Model, The training loss keeps decreasing, however, the test and the validation loss converge in around 5-7 epochs. The model shows signs of over fitting as the test accuracy increases slightly afterwards on further training.

6.1 Example generated text

Epoch 0 (male, pop rap): I ain’t in nothin’ else, I’m sorry but still in yet I don’t live it on regrets, And I just used to say I hate him in dishonest jest

7. CONCLUSION

In general, training Generative Adversarial Networks (GANs) is a hard problem for textual input, because text is inherently discrete, and therefore we cannot take perform mini updates in any direction for the model to learn,
the model still needs more improvement.

8. REFERENCES

