Lyrics Generator

Jun Wang, Pengsen Li, Junke Wang, Hanlong Peng
Georgetown University ANLY 590 Neural Nets and Deep Learning

ABSTRACT

Music creation performs multiple roles in people's life. It serves as art creation, therapy process, communication media, and many other important functions. We present a lyrics generation method with deep learning knowledge to construct a meaningful, interesting lyric story with specific singer's style. The method is based on a character-level RNN with a hidden GRU model as well as a spelling checker function and a timer. We first construct three different numbers of layer for the GRU model (1-layer, 2-layers, and 3-layers) and compare both word-level and char-level accuracy. Results show that 2-layers GRU model provides the highest accuracy percentages - 95.48% for word-level and 90.97% for char-level. Second, we try various hidden sizes for the 2-layers GRU model (size of 5, 20, 50, 100, and 200). As a result, a hidden size of 100 and 200 have relatively high accuracy percentages (above 95% for the word-level and above 91% for the char-level. Ultimately, the model generates an interesting and mimic short lyric paragraph.

INTRODUCTION AND PROBLEM STATEMENT

Music creation can be tricky since it requires not only the theories from books and experiences from life but also writers creativity and inspiration. Although sophisticated technologies and software are developed for songwriting, it is still hardly possible for any singer or writer to produce great works with a similar style in a frequent and constant rate. The lyrics generator in this research uses the existing lyrics from the songs of two famous singers as its database. We believe that this generator can assist any user to create complete lyrics for a song regardless of talent or creativity.

RELATED PREVIOUS WORK

Among various related approaches for lyrics generator problem, the most popular and efficient methods are Hidden Markov Models and Recurrent Neural Network.

- Hidden Markov Models
  - Hong Kong University of Science and Technology
  - Unsupervised Rhyme Scheme Identification in Hip Hop Lyrics Using Hidden Markov Models
  - Obtain a precision of 57.25% on identifying the rhyming words
  - With a total F-Score of 44.06%
- Recurrent Neural Network
  - Finland
  - Method based on the RankSVM algorithm and a deep neural network model
  - Model outperforms the best human rappers by 21%
  - Deployed as an online tool called DeepBeat

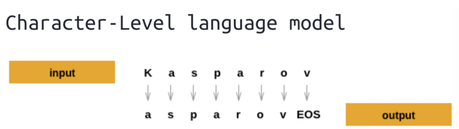
DATASETS

Eminem, one of the best rappers of all time, is selected to be the singer we obtain the training data from.

- Eminem Lyrics Dataset
  - Web scraping from https://www.azlyrics.com/ (beautifulsoup-python)
  - Song title; Lyrics
  - 350 songs

METHODOLOGY

The basic model for our lyrics generation task is Char-RNN., which is a recurrent neural network trained to predict the next character given a sequence of precious characters.

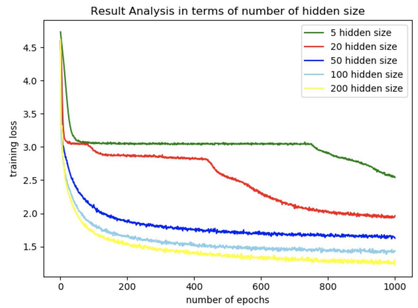


This could be also explained as a conditional probability distribution to find the most likely character from the alphabet including some other signs. Note that we do not need to tokenize before we define the dictionary of our character alphabet since the model is in character level. When creating training sequences, we set the chunk length of the seed characters equals 200 and each character is typically encoded as a one-hot vector.

In practice, the main RNN model we choose is GRU model, which is faster trained than LSTM. We try three different numbers of layers (1-layer, 2-layers, and 3-layers) with the temperature equals 0.7, the hidden size of 100, training epochs of 1000, and batch size of 256. The following table is the word-level and char-level accuracies with time results for the three GRU models:

number of GRU layers	word_acc	char_acc	Time
1	93.73%	87.99%	0:23:21
2	95.48%	90.97%	0:26:50
3	94.52%	89.15%	0:36:55

Since the 2-layers GRU model provides the best accuracy results, we decide to use it as the basic model for the next step. With all parameters fixed in the 2-layers GRU, we change the value for the hidden size for the model. The losses for each epochs are shown below:



Since these are the losses for the training set, it make sense that the larger hidden size is, the better the model performs. So we take a look at the accuracies and time for each model:

hidden size of GRU	word_acc	char_acc	Time
5	71.295%	56.615%	0:28:19
20	86.344%	75.578%	0:28:24
50	92.145%	85.160%	0:26:41
100	95.555%	91.467%	0:26:36
200	96.339%	92.254%	0:28:27

From the above table, we can see that hidden size of 100 and 200 both have high accuracies of word-level (above 95%) and char-level(above 91%). Therefore, we choose these two models to generate the Eminem-like Lyrics with the say starting word "Hey".

ANALYSIS OF RESULTS

Lyrics output from the model with hidden size of 100:

Hey
you money to beful
i shit i don't got a little back, i hear the stupid
so mad you should tell it and i do it
imma be really faking in the honesher doubf
custom what you know you wanna man through the peach
but i can screamings around it, so they ha would it gonna see me
and i got down, down it the past ruthing they talk it that's every gone
what you let's beginning a hole it, i feel hours againma some days and vister
and i up and let pocket your channipositions that with millions
so i can 'neal

Lyrics output from the model with hidden size of 200:

Hey
but i just don't give a fuck
i wanna take me too, she has told you shot
'cause you give'em the new rist so we can looks like a fuckin' colla
don't want you no fame, want you no gut
so they say they're just a chuck
and when i go to you, but it's just the type of my back
slam with me, let on a soldier who wanna rock the heck
when i just choke of it, it subla
cause i smiles and shook this head while we going to rap
when i'm slim shady
so you just get one control
come on a second shot
we came off the flow

By comparing the above outputs, it is obvious that lyrics from the 200-hidden-layer model is more fluent and natural (with accuracy of 89.05%). There are misspelling words like 'beful' and 'visiter' in the 100-hidden-layer model and the rhyme is worse (with accuracy of 82.77%).

REFERENCES

[1]. Unsupervised Rhyme Scheme Identification in Hip Hop Lyrics Using Hidden Markov Models: https://link.springer.com/chapter/10.1007%2F978-3-642-39593-2_3
[2.]. DopeLearning: A Computational Approach to Rap Lyrics Generation: https://dl.acm.org/citation.cfm?id=2939679&dl=ACM&coll=DL
[3].https://datascience.stackexchange.com/questions/13138/what-is-the-difference-between-word-based-and-char-based-text-generation-rnns
[4.]. https://github.com/spro/char-rnn.pytorch/blob/master/train.py
[5.]. https://github.com/minimaxir/textgenrnn