COL 772 : Course Project Proposal

Rap lyric generation using LSTM

Team 1
Harish Chandra Thuwal (2017MCS2074)
Deepak Sharma (2017MCS2071)

Aim :
The aim of this project is to generate unconstrained rap lyrics pertaining to an artist using LSTM based architecture.

Data sets :
We would use the freely available lyrics archive *OHHLA* [1], which has lyrics available for almost all songs/artists in plain text formats. We would develop a simple web crawler to obtain the required texts and parse them into suitable format.

Literature review:
Peter Potash, Alexey Romanov, Anna Rumshisky , in their paper *"GhostWriter: Using an LSTM for Automatic Rap Lyric Generation"* [2], have used an LSTM based system for generating novel lyrics that are similar in style to a target artist. Their system improves upon the n-gram model by (Barbieri et al ,2012).

The code for their system is directly available at github[3]. They have used theano to implement the system. We intend to replicate their reported results and then improve upon further.

Evaluation :
For evaluation, we would use the metrics (as used in referenced paper) of rhyme density and max similarity. Low max similarity and high rhyme density will be an indication of novelty of lyrics and similarity to artist style respectively. Since exact replication would assuredly give a higher rhyme density than randomly produced lyrics, it's also desirable to have a low correlation between rhyme density and max similarity.

Also, similarity will be used as a performance metric, when system generates lyrics that have rhyme density similar to the given average rhyme density of the target artist. Further we would aim to incorporate other evaluations methods put forward by the authors in [4].

Demo :
We intend to have a simple demonstration wherein the user will have the options of different artists for generating sample lyrics , with a seed verse/phrase to start with.
cd
References :
[1] The Original Hip-Hop (Rap) Lyrics Archive - OHHLA.com - Hip-Hop Since 1992 .

[2] Peter Potash, Alexey Romanov,and Anna Rumshisky.2015.Ghostwriter: Using an LSTM for automatic rap lyric generation. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing.
[3] https://github.com/jgc128/Ghostwriter
[4] P Potash, A Romanov, A Rumshisky - Evaluating Creative Language Generation: The Case of Rap Lyric Ghostwriting . arXiv:1612.03205, 2016