

Teddy Morris-Knowler
Santiago Seira

2Pac Resurrected: a Machine Learning Approach

Define input-output behavior:

For each of 5 rap artists (2pac, Notorious B.I.G, Kanye West, Jay-Z, Kendrick Lamar), we will take 50-100 of their songs and compile all of the lyrics together for each rapper. We will train our model on these lyrics for each of the rappers, and the output will be a “freestyle,” or a new verse by the rapper in their given style.

Define evaluation metric for success:

Touring test - get a human that doesn't know the artist, give them 3 verses, 2 real 1 generated, and ask them to determine which one is which. If > 50 percent success rate (meaning choose real verse as the fake one), then successful.

Oracle:

Creates songs that no one can distinguish from real artist, even people who know the artist well can't tell which is the fake one. Be able to sell the rights to the artist itself, create “ghost writers,” which generate lyrics for artists.

Baseline:

Most baseline: random shuffling of words artist uses. 2nd baseline: ngrams model from 106b where generate map of words to words that come after those words in the artists lyrics.

Gap:

Finding a way to create original music, incorporate words they haven't explicitly strung together, but still making it sound like their “style.” Involving creation of lyrics, not just shuffling up the phrases they have used.

Challenges:

Come up with original work, not just repeating words that the rappers have said, but thinking about what words they *would* say and what phrases they *would* use if coming out with a new song.

Create lyrics with literary tactics

- Metaphors

- Symbolism

- Assonance

- Alliteration

Match rhyming scheme, also understanding how accents + style make words rhyme when they shouldn't: (“You know it's hard out here for a pimp / when you tryna get that money for the rent”)

Motifs within a song, getting theme to play out throughout the song (“Hustler” story, fame story,

Matching their words to already existing beats: making it “sound like rap,” not just words that they would say.

Scraping HTML to get lyrics from rappers and

Topics:

Clustering of raps within a rapper to distinguish ‘style of rapper’

Train on data: feature extraction + develop weight vector

Search problem: start w/ a path of no words, continue to iterate until cost is minimized between path (generated rap) and training data. Utilizing n -gram model + adding.

Similar work:

Fluency: a poetry generating algorithm that takes Shakespeare's sonnets and creates new, original word based on them.

<http://www.dailymail.co.uk/sciencetech/article-2546811/The-robot-writes-POETRY-Machine-help-s-compose-sonnets-studying-Shakespeares-work.html>

DeepBeat, a machine learning algorithm that both generates raps based on over 100,000 songs from 100 different rappers, and also is able to evaluate rappers based on their rapping density and can rank them in accordance to lyrical complexity.

<http://www.technologyreview.com/view/537716/machine-learning-algorithm-mines-rap-lyrics-the-n-writes-its-own/>

<http://arxiv.org/abs/1505.04771>

Data:

Will scrape:

<http://www.allthelyrics.com/lyrics/2pac>

http://www.allthelyrics.com/lyrics/notorious_big

<http://www.allthelyrics.com/lyrics/drake>

http://www.allthelyrics.com/lyrics/kanye_west

<http://www.allthelyrics.com/lyrics/jayz>