

Repetition of Song Lyrics & Music Genre Classification

Wenjie Xie

2/5/2020



According to the psychologists, human is such a sucker of repetition. Our brains are more willing to process the information that we are already familiar with. That's why, in the music industry, song producers have to admit that repetition is the central element of music, especially in the modern music industry. In this report, let's go and explore the interesting facts about repetitiveness of song lyrics.

Basic Data Cleaning

In order to compare the repetitiveness of lyrics, I measure the length of unique words in each song and calculate the drop rate of lyrics size, which I called "reduction rate" in this report.

Ariana Grande, problem

one less, one less problem

one less, one less problem

head in the clouds

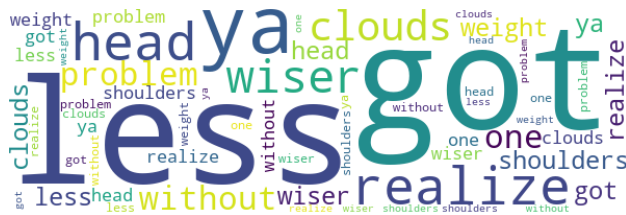
got no weight on my shoulders

i should be wiser

and realize that i've got (i've got)

one less problem without ya!

one less problem without ya!

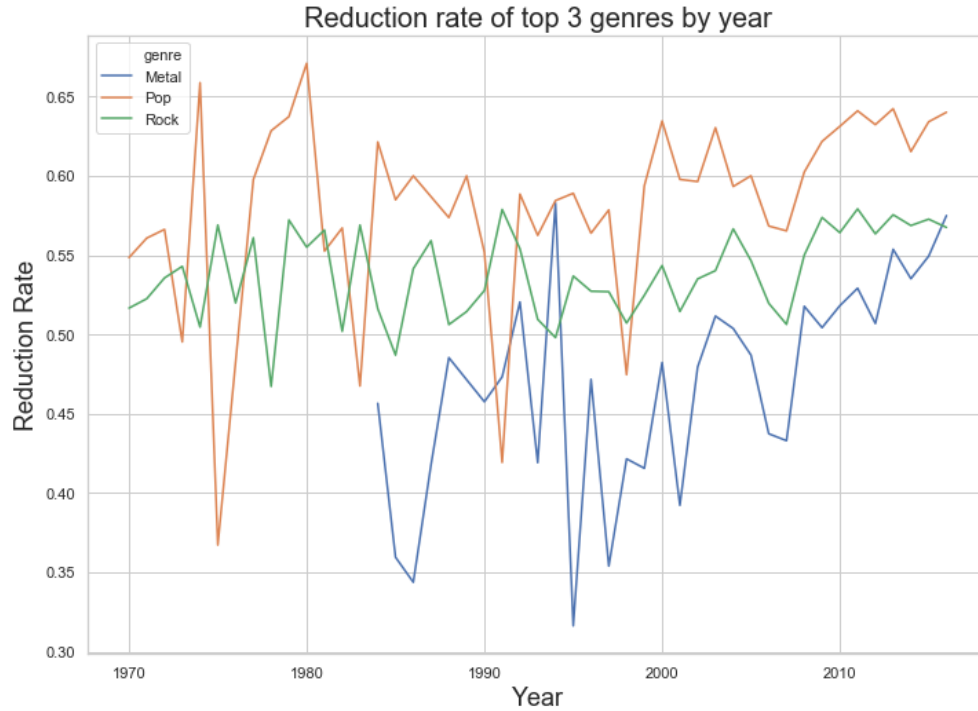


The line graph illustrates the annual reduction rate of song lyrics from 1970 to 2015. The y-axis, labeled 'Reduction Rate', ranges from 0.50 to 0.58. The x-axis, labeled 'Year', spans from 1970 to 2010, with data extending to 2015. The graph shows a highly volatile trend with several peaks and troughs. A significant peak occurs around 1978, reaching approximately 0.585. Another major peak is seen around 2012, reaching approximately 0.588. The rate generally fluctuates between 0.51 and 0.55 for most of the period, with a notable dip around 1976 and 1998.

Year	Reduction Rate
1970	0.524
1971	0.550
1972	0.548
1973	0.540
1974	0.510
1975	0.542
1976	0.518
1977	0.565
1978	0.585
1979	0.578
1980	0.558
1981	0.520
1982	0.540
1983	0.545
1984	0.532
1985	0.525
1986	0.530
1987	0.528
1988	0.515
1989	0.528
1990	0.535
1991	0.575
1992	0.550
1993	0.515
1994	0.532
1995	0.535
1996	0.532
1997	0.528
1998	0.508
1999	0.538
2000	0.548
2001	0.530
2002	0.545
2003	0.555
2004	0.560
2005	0.555
2006	0.525
2007	0.502
2008	0.565
2009	0.578
2010	0.575
2011	0.588
2012	0.582
2013	0.585
2014	0.578
2015	0.585

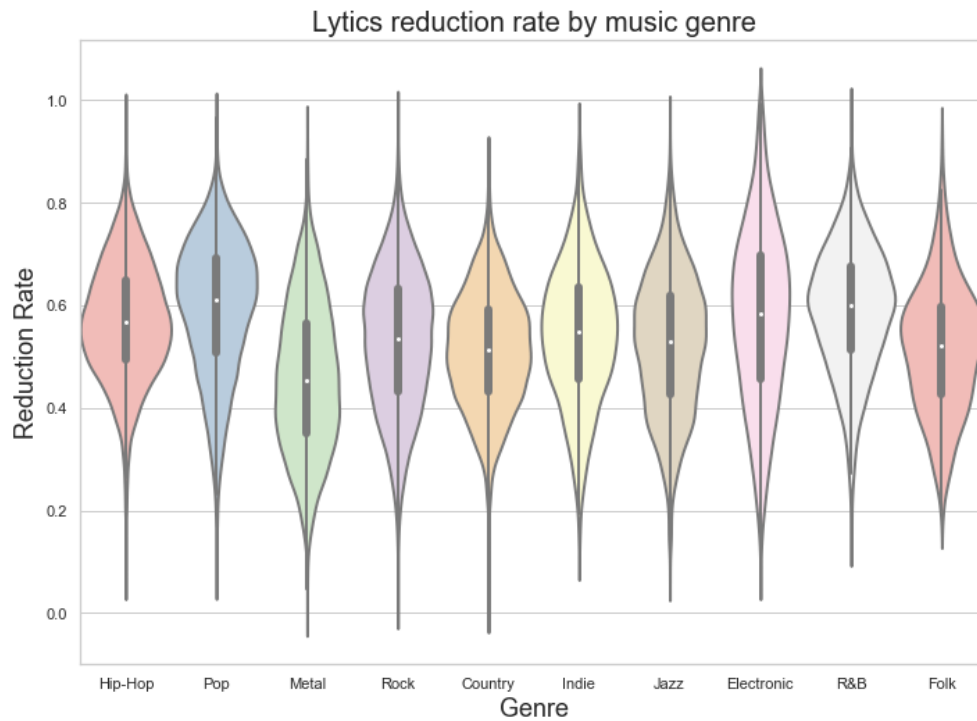
2

What if we look closely to the trend of repetition in some specific genres?



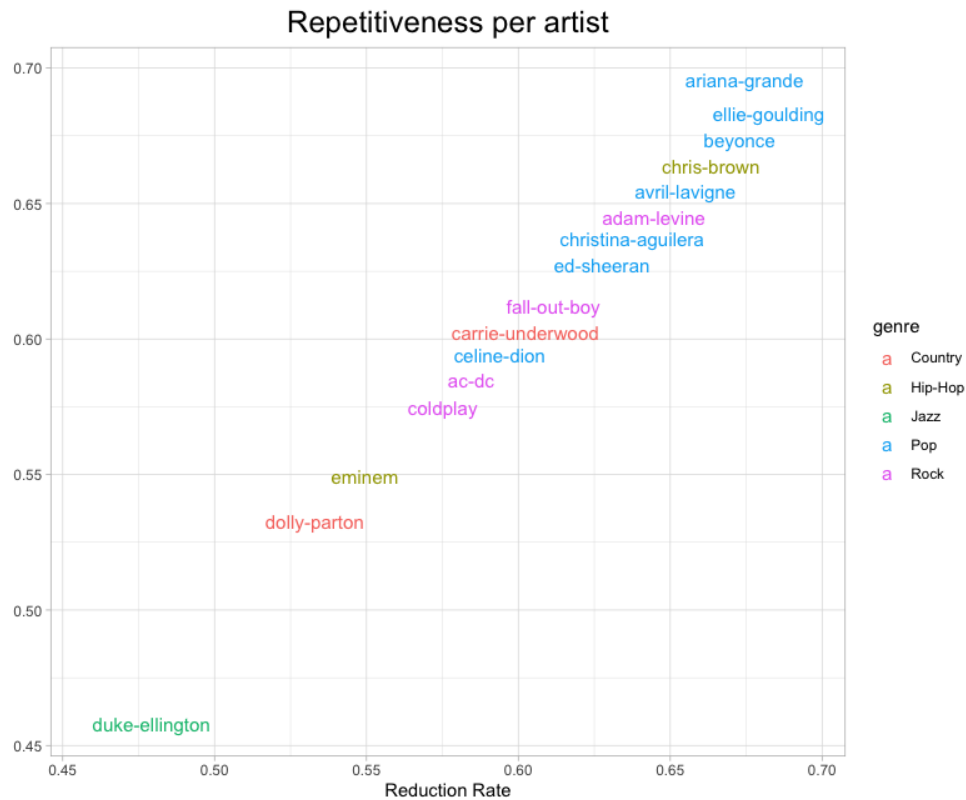
In this step, I look through the top 3 genres with the greatest number of songs included in this dataset, which are Rock, Pop and Metal music. Due to the incompleteness of the dataset, in 1975, only several songs of a single artist, Gino Vannelli exists which are classified as pop music. Except these abnormal reduction rates, we could generally see the noticeable differences of the average reduction rate between 3 genres: Pop higher than Rock, and Metal always be the lowest.

Lyrics reduction rate by music genre



Based on this finding, I started to explore the repetition characteristics from all music genres. From this violin plot, we could observe that on average, Pop music has the most repetitive vocabularies and Metal get the least. It is quite easy to interpret in the way that in order to make the song appealing or catch listeners' ears, the catchy hook with repetitive sentence are necessary in music genres like Hip-hop, Rock, R&B, and Pop, which results in the high reduction rate. On the other hand, the styles of Metal or Jazz are less predictable. From melody perspective, Metal is more chaos, and the lyrics are more critical, which requires more vocabularies to convey the topics.

Repetitiveness per artist



In this graph, I plotted the average reduction rate of selected known artists. The more top right the artists be, the higher their reduction rates are. Most of the artists' discographies conform to the characteristics of lyrics repetition rate of each genre. Pop music artists, with high reduction rate, gather on the top right corner; Duke Ellington, representing jazz, and country music ambassador Dolly Parton are positioning on the bottom left.

One of the interesting findings is that with even with reduction rate of 0.64, Adam Levine's songs are still considered as Rock. Diving deep into the data, I spotted "Lost Star", which is a really repetitive song. According to Wiki, this is absolutely a pop music. Adam Levine is the lead vocalist of rock band Maroon 5, which might be the main reason of sorting his music as rock. But I think the identity of the singer shouldn't be the only criterion to classify the works. Based on this finding, I am wondering whether repetitiveness would help the music industry to improve music genre classification?

Summary

By analyzing the repetitiveness of song lyrics, we could get the following results.

- On average, pop music is most likely to contain repetitive vocabularies and sentences in the lyrics. Other music genres like Jazz, Country and Metal normally have lower repetitiveness.
- We know the music classification should be based on both audio and lyrics. In the study of lyrics repetitiveness, we could see the potential of lyrics contribution to music classification even from very shallow textual features. With further study that will reach the semantics level of analysis, we could gain more information that benefits music recommendation system.

Reference

1. Mayo Oshin. The science of how music affects your productivity. <https://thriveglobal.com/stories/why-music-can-make-you-more-productive/>
2. Colin Morris. Are Pop Lyrics Getting More Repetitive? <https://pudding.cool/2017/05/song-repetition/>
3. Michael Fell, Caroline Sporleder. Lyrics-based Analysis and Classification of Music.