Lyrical Trends in Hip-Hip and R&B since 1975

Rohan Pahwa and Eliot Kim

Abstract

This study will analyze patterns in songs in hip hop based on a five-year interval starting from 1975. Using n-grams, the change in relative frequency distributions will tell us how much certain words or phrases have become popularized throughout time. We will also look at some major global events and see if the lyrics during the time are positive or negative depending on the global event. We hope to find trends in lyrics and know when if not how certain phrases or words came into our vocabulary as common words.

1 Introduction

We chose to analyze song lyrics in between a certain period of time to see if we can analyze patterns or words that change over time. For instance, we might be able to find specific words in which did not appear in a lot of songs earlier on but became popular as time went on. We can also look for patterns around landmark events to see if that had an impact on the type of words that were used in lyrics. This could be interesting because these kinds of patterns can give us a new perspective on our social history and understand how the music industry played a role in contributing to its growth. It is almost a different way of seeing a historical timeline in the sense that we will be able to hopefully identify familiar trends and themes that can give us a better idea on the origins of some words and patterns. We believe that we would be surprised to see how far back or how recently some words we use commonly today first appeared in songs or the themes we see in the songs we enjoy based on the time it was written.

2 Background

A study conducted by Smith, Zee, and Uitdenbogerd analyzed music lyrics to find trends of cliché lyrics and themes in all music. Lyrics were pulled from LyricWiki using a web crawler. Lyrics were then checked to see if duplicates were found to prevent double analysis. Method of analysis like rhyme-pair, lyrical attributes, and collocations (trigrams) were used to find repetition and cliché themes in the music. Smith et al. (2012) found that after taking the Billboard top 100 from the time period and applying the testing trained using the LyricWiki data they were able to distinguish from highly cliché and slightly cliché. They struggled with music that was closer to the middle of the pack in terms of cliché-ness. I'm concerned to see how they developed a baseline of clichés. It seems like they used subjective values to define a baseline.

A study conducted by Fell and Sporleder is very similar to the goals that we want to achieve with our project. The goal was to analyze lyrics for genre detection, distinguishing the best and the worst songs, and determining the approximate publication time of a song. They did this by using n-grams and certain features such as style, semantics, length, rhyme and vocabulary to achieve these 3 goals. Their results were promising but difficult to assess because of their inability to compare their results since their methods were rather new. They mentioned that adding additional features always improves the results. A concern would be that as you add more features there is a possibility of losing accuracy in identifying these features and having too many features could make the results too exclusive to the features added. Finding the optimal number of features rather than the most would be the challenge.

Patra, Das, and Bandyopadhyay conducted a study in which analyzes the mood in Hindi songs. Their goal was to analyze lyrics and text to determine positive or negative polarity. They did this through many different methods, including a mood taxonomy, analyzing the audio, and stylistic features. This study is relevant to our study because we hope to recognize a pattern in which songs were produced around global events, positive and negative and see how that affects the songs produced at that time. Their results were conclusive with the main issues being the range of their textual features and the contradiction between the audio and lyrics. Independently, analyzing audio and lyrics could be accurate but trying to incorporate the results of both can lead to a lower accuracy of determining the mood of the song. Knowing which features to incorporate and how to implement them without it lowering the accuracy is a concern for their study.

3 Data

A large corpus of R&B and Hip-Hop Lyrics was not available so we created our using a few different tools. A list of the top five songs by week from the week of May 5th, 2018 till the first week of 1975 was scraped from the https://www.billboard.com/charts/r-b-hip-hop-songs/ website. We used the billboard-api develeoped by Allen Guo to achieve this. This list was exported to a .csv file with 11,310 entries. We then removed all duplicates and fixed all the broken entries. Then we used another the lyricwikia API by Enrico Bacsis and downloaded all the lyrics from http://lyrics.wikia.com/wiki/LyricWiki site. Once we downloaded all the lyrics we combined them into nine groups by half decade going from 1975 to 2018. These groups were converted into single files set up for processing by our code. We made sure to remove artist tags and any other irrelevant information. Duplicate songs were removed because they only show that the song was popular for multiple weeks. We were more interested in the lyrical trends rather than the popularity of the songs. This caused a data size discrepancies between the half decades but this was dealt with using relative frequencies instead of raw frequencies when looking at the unigrams, bigrams, etc.

Time	Number of Documents	Number of Sentences	Number of Words
1975	244	11,854	64,386
1980	210	11,865	61,659
1985	336	19.993	103,370
1990	201	13,137	71,730
1995	144	9,764	57,058
2000	169	13,485	81,572
2005	162	14,470	82,167
2010	146	11,312	63,580

Table 1: Data

1975			1980		
Word	Freq	Relative Freq	Word	Freq	Relative Freq
love	977	0.015174106	love	634	0.010282359
baby	352	0.005467027	know	423	0.006860312
like	346	0.005373839	night	325	0.005270926
got	334	0.005187463	got	294	0.00476816
know	329	0.005109806	time	272	0.004411359
yeah	289	0.004488553	like	271	0.004395141
get	275	0.004271115	want	254	0.004119431
Oh	249	0.0038673	yeah	252	0.004086995
night	244	0.003789644	baby	235	0.003811285
way	235	0.003649862	one	234	0.003795066
want	225	0.003494549	Oh	233	0.003778848
good	216	0.003354767	say	228	0.003697757
oh	203	0.003152859	gon	209	0.003389611
time	187	0.002904358	never	186	0.003016591
We	182	0.002826701	get	184	0.002984155
stop	182	0.002826701	let	178	0.002886845
see	179	0.002780107	heart	177	0.002870627
When	177	0.002749045	We	173	0.002805754
come	175	0.002717982	Let	172	0.002789536
right	173	0.00268692	make	171	0.002773318
64,386			61,659		

Table 2: Results

1985			1990		
Word	Freq	Relative Freq	Word	Freq	Relative Freq
love	1430	0.013833801	love	772	0.01076258
know	788	0.007623101	baby	514	0.00716576
oh	702	0.006791139	know	486	0.00677541
baby	615	0.005949502	got	339	0.00472606
got	557	0.005388411	like	337	0.00469817
Oh	468	0.004527426	make	310	0.00432176
want	463	0.004479056	get	308	0.00429388
We	407	0.003937313	time	291	0.00405688
yeah	403	0.003898617	want	276	0.00384776
time	383	0.003705137	go	270	0.00376412
like	365	0.003531005	yeah	268	0.00373623
make	350	0.003385895	way	261	0.00363865
heart	350	0.003385895	heart	243	0.0033877
go	349	0.003376221	see	242	0.00337376
gon	344	0.003327851	never	235	0.00327618
feel	343	0.003318177	one	230	0.00320647
never	338	0.003269807	gon	229	0.00319253
one	333	0.003221438	man	217	0.00302523
need	329	0.003182742	Oh	216	0.00301129
get	300	0.002902196	feel	216	0.00301129
103,370			71,730		

Table 3: Results

1995			2000		
Word	Freq	Relative Frec	Word	Freq	Relative Frec
love	539	0.00944653	know	595	0.00729417
know	475	0.00832486	like	543	0.0066567
baby	396	0.00694031	love	516	0.0063257
like	347	0.00608153	got	488	0.00598245
got	299	0.00524028	baby	413	0.00506301
make	272	0.00476708	get	397	0.00486687
yeah	266	0.00466192	yeah	324	0.00397195
want	246	0.0043114	girl	321	0.00393517
oh	236	0.00413614	go	311	0.00381258
one	208	0.00364541	right	268	0.00328544
go	194	0.00340005	Oh	267	0.00327318
see	194	0.00340005	gon	264	0.00323641
get	188	0.00329489	make	251	0.00307704
need	178	0.00311963	need	246	0.00301574
gon	176	0.00308458	back	243	0.00297896
body	175	0.00306706	oh	241	0.00295445
never	165	0.00289179	'Cause	239	0.00292993
back	159	0.00278664	see	238	0.00291767
Oh	155	0.00271653	way	238	0.00291767
come	150	0.0026289	say	235	0.00288089
57,058			81,572		

Table 4: Results

2005			2010		
Word	Freq	Relative Frec	Word	Freq	Relative Frec
like	857	0.01042998	like	598	0.00940547
got	678	0.00825149	oh	587	0.00923246
know	672	0.00817847	love	463	0.00728216
get	469	0.00570789	know	382	0.00600818
love	452	0.00550099	yeah	317	0.00498585
girl	340	0.00413791	get	307	0.00482856
see	330	0.00401621	baby	307	0.00482856
go	303	0.00368761	got	302	0.00474992
So	292	0.00355374	We	277	0.00435672
want	290	0.0035294	want	248	0.0039006
gon	278	0.00338335	Oh	245	0.00385341
one	274	0.00333467	gon	232	0.00364895
yeah	266	0.00323731	go	231	0.00363322
baby	264	0.00321297	never	230	0.00361749
make	258	0.00313995	let	222	0.00349166
way	255	0.00310344	night	215	0.00338157
ai	251	0.00305475	'Cause	212	0.00333438
let	238	0.00289654	one	203	0.00319283
oh	238	0.00289654	make	189	0.00297263
ya	235	0.00286003	hey	186	0.00292545
82,167			63,580		

Table 5: Results

2015		
Word	Freq	Relative Freq
like	346	0.00818857
got	306	0.00724192
know	304	0.00719459
love	266	0.00629526
oh	231	0.00546694
let	182	0.00430729
one	179	0.00423629
work	175	0.00414162
go	158	0.00373929
get	153	0.00362096
baby	150	0.00354996
yeah	149	0.00352629
Oh	145	0.00343163
We	145	0.00343163
feel	143	0.0033843
need	137	0.0032423
time	127	0.00300563
want	117	0.00276897
So	109	0.00257964
'Cause	107	0.00253231
42,254		

Table 6: Results

4 Method

We primarily computed most of our results in python using the nltk library. After we clustered the text files of lyrics based on the year, we simply read in these files in python, removed the new lines, and tokenized them. We then implemented the stop word corpus from nltk and added additional basic stop words based on what we have seen so far. After removing these stop words from the tokenized lyrics, we were able to create n-grams and run frequency distributions.

We ran multiple experiments depending on the n-gram we decided to run. We had to add additional stop words depending on the n-gram to make sure that the top twenty results from the frequency distribution were relevant and useful. We also had to experiment with how many n-grams we should run in terms of unigrams, bigrams, trigrams, etc. We decided to go up to 5-grams because of some of the repetition in words that were shown in n-grams up to 4-grams. We took these n-grams and calculated the relative frequencies for the top 20 unigrams and bigrams. This gave us a clearer understanding of word choice in comparison to the raw frequency values provided by nltk.

5 Results

The unigrams provided the most useful information to us. We were able to notice some basic trends about the shift in vocabulary. We noticed from the baseline that some words that we expected to see in a list of hip-hop songs were simply not there. Instead the most popular words were ‘love’, ‘like’, and ‘baby’. One of our goals was to find themes in the lyrics that might have to so with the social mentality of the time, we thought we might see race or socioeconomic based lyrics be on top but neither really showed up. A lot of what we expected to see, didn’t actually show up. This is because of the quantity of words that we had in our data set and our inability to identity phrases because of constant repetition.

We were also able to conclude that while we calculated up to 5-grams, most of the words that showed up in the frequency distributions were the same because of the word repetition. We originally thought that increasing the n-gram would result in more phrases or sentences, but instead, the results showed the same word simply repeated multiple times. As a result, even though we calculated multiple n-grams, it was difficult to find trends in phrases as we originally planned, but rather found more trends in words. However, the repetition of words provided us with an unexpected finding. We were able to identify changes in how more recent songs had a tendency to repeat words than that of older songs. We were also able to see, as previously mentioned, the transition of words from “like” to “love” throughout time.

6 Conclusions and Future Work

We learned that the variety of words and the number of repetition has significantly changed over the years. Words that don't carry much meaning are repeated more in recent songs than that of songs from other decades. Overall we noticed a sharp uptick in the sheer quantity of repetition on lyrics. Songs like Rihanna's "Work" and Lil' Pump's "Gucci Gang" took over the frequency distributions as soon as we began looking at anything greater than a bigram. Even half decades before the current had lyrical repetition but never to the level of modern artists. This trend seems to have started the post 2000s

We succeeded in organizing our data in a way that is efficient and easy to understand. Clustering our data into intervals based on years and storing the corresponding lyrics based on the year made it extremely easy when tokenizing when we were experimenting creating the n-grams. When running experiments, it was easy to understand what data result came from what year and learn from it.

Where we failed was the strength of our results. While we were able to identify some trends and findings, it was difficult to identify phrases or words that we didn't know were popularized in previous years. Simply the number of irrelevant words and variety of them made it difficult to add or remove certain classifiers to narrow down our findings.

Ways to make our results more accurate and meaningful is the data that we use and how we analyze it. For instance, we focused on corpus of lyrics on hip hop and R&B because we assumed that these types of songs typically contain phrases or words that become popular in society. However, while we were able to identify common words, it was difficult to identify a trend of phrases because of the amount of repetition in the use of words. Not only that but phrases or words can become popularized or well-known through a single song, which made it difficult to pinpoint phrases. Therefore, if we had a database in which we had access to popular phrases or words, we can use this information to compare when and then how often these phrases or words were used from that point forward.

In addition, in order to take care of the issue that we had with the repetition of words, setting a limit to how many words can be repeated could be valuable. This way, it could potentially show more phrases or sentences when we conduct 5-grams or more instead of words being repeated over and over again. This could potentially help in finding successful results in what we were originally attempting to find.

What we should have done differently is to have trends or specific words to look for to make our results more meaningful. If we had a source in which we could look for a specific theme or trend based on that time period, we could've added another classifier or additional stop words in which filtered out more unnecessary words and isolated the n-grams in which gave us valuable data to better see the development of slang and phrases.

References

Smith, Alex G, et al. "In Your Eyes: Identifying Clichés in Song Lyrics." Proceedings of the Australasian Language Technology Association Workshop 2012, Dec. 2012. ACL Anthology, aclanthology.coli.uni-saarland.de/papers/U12-1012/u12-1012.

Fell, F. Sporleder, C. 2014. COLING. Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. August 2014. Pp. 620–631.

Dipankar Das, and Sivaji Bandyopadhyay. Mood Classification of Hindi Songs based on Lyrics Braja Gopal Patra. Department of Computer Science & Engineering