I've Got a Feeling: Analyzing the Lyrics of Popular Music in the United States During the British Invasion Era

Chongyang Chen   Dongjie Fan

cc5502@nyu.edu   df1676@nyu.edu

June 11, 2017

Abstract

British Invasion, a musical movement in the mid-1960s, a time in which numerous British artists expanded their popularity to the United States. In this paper, we are interested in studying the commonalities and differences between lyrics of songs performed by the British Invasion artists and other artists who had appeared on the Billboard Hot 100 charts during that period of time. Result of the LDA topic model suggests a subtle difference between the lyrics of songs performed by British Artists and their American peers.

Background

In the music history of the United States, British Invasion refers to a cultural phenomenon that took place during the 1960s. On February 9th, 1964, The Beatles made their first American television appearance on The Ed Sullivan Show and had quickly become the most popular artists in America. Following the footsteps of the Beatles, many British

---

bands, such as the Rolling Stones, the Who, the Kinks, the Animals, and Herman's Hermits, entered and succeeded in the United States. While British Invasion drastically influenced the landscape of American music industry and broadened the scope of pop music, in the 1960s, the names of Elvis Presley, Ray Charles, Dionne Warwick, and Frank Sinatra could still be seen on the top half of Billboard Hot 100 charts; the likes of Bob Dylan, the Beach Boys, and Marvin Gaye were showcasing their talents in different styles and genres.

In this project, we collected the lyrics of songs that had made at least one appearance on the Billboard Hot 100 weekly charts from the first week of 1964 to the last week of 1969. We are interested in studying the commonalities and differences between lyrics of songs performed by the British Invasion artists and other artists who had made the Hot 100 charts.

Theory and Hypothesis

English bands like the Beatles and the Rolling Stones were influenced by the American rhythm and blues (R&B), or Rock 'n' Roll music in the 1950s. Since some of the Rock 'n' Roll artists that were popular during the 50s were still active in the 60s, and the British Invasion bands had covered a considerable amount of R&B songs (the Beatles: "Twist and Shout" (by the Isley Brothers), "Johny B. Goode" and "Roll Over Beethoven" (Chuck Berry), "Boys" and "Baby it's You" (the Shirelles); the Rolling Stones: "Beautiful Delilah" and "Bye Bye Johnny" (Chuck Berry), "I Can't be Satisfied" and "I Want to be Loved" (Muddy Water) etc.), we believe that the lyrical styles of British artists would not be significantly different from their non-British counterparts.

As Gillet argues, "the Beatles ... did not invent a style but simply drew attention to sounds and styles that were current in the United States", [3] we believe that songs performed by British artists during the British Invasion era were not fundamentally different

from the popular songs produced by their American peers. The popularity of the Beatles, and other ensuing British bands in the United States during the 1960s is a result of a combination of personal charisma, successful business and marketing operations, vast amount of target audiences (World War II baby boomers), and performance styles including their hair style, the way they dress up, and their accents.

Our hypothesis is that there might be subtle difference between the lyrics of songs performed by the British Invasion artists and their peers on the charts. We are planning on attempting to interpret the subtle difference between these two groups using text mining techniques such as sentiment analysis and topic modeling, although statistically significant findings are more than welcome.

Literature Review

Few researches have been conducted to analyze lyrics. Ransom explored the relationship between lyrics and positive psychology [4].Burns took out the top 20 popular songs from 1963 to 1972 and labeled their themes based on his qualitative analysis. [2] Neither of the articles propose a generalized method that is scalable in analyzing the sentiment of lyrics and modeling the topics, messages, or themes each song conveys.

Data and Preprocessing

Data

Our data is comprised of 313 Billboard Hot 100 Weekly charts from the first week of 1964 to the last week of 1969. Based off the names of artists and tracks on the data, we attempted to collect lyrics data from sources like lyricsfreak.com and azlyrics.com. The "List of British Invasion Artists" page on Wikipedia provides the list of artists that are commonly identified as members of the British Invasion, and we marked the songs performed by these artists as "BI" ("British Invasion"). The other songs in the data

are marked as "AM", which stands for "American Peers", though not all of the these artists are American (the Bee Gees, for instance). We further removed two songs that are written in non-English languages.

	Songs (count)	Songs (%)	Artists (count)	Artists (%)
British Invasion	280	25.41	36	31.03
American Peers	822	74.59	80	68.97
Total	1102	100	116	100

Preprocessing

When creating the document-feature matrix (dfm) from the corpus, punctuations, numbers, symbols, hyphens, and a shorter version of quanteda package's default English stop words were removed. The terms we chose to keep from being removed are singular personal pronouns such as "I", "you", "he", and "she". In addition, several common interjections, like "yeah", "na", "la", "whoa", etc. were also removed. In order to keep each entry non-zero, the default smoothing function was applied to the matrix. Finally, in trimming the matrix, feature with less than 20 occurrences (1122 after smoothing) were removed from the dfm to enhance computation efficiency.

Result

Similarity

The "impact" of a particular artist is measured by the number of times the artist's song, or songs appear on the Billboard Hot 100 weekly charts A. A subset of the corpus was created to conduct a hierarchical clustering on the euclidean distance matrix of the term frequency–inverse document frequency (tf-idf) measure of selected artists that had a higher than 15 "impact". 7 of the 39 artists were British Invasion artists (marked in green in Figure 1).

The result of hierarchical clustering (Figure 1) shows that, although there is no clear evidence to prove that lyrics of songs performed by the British artists are collectively

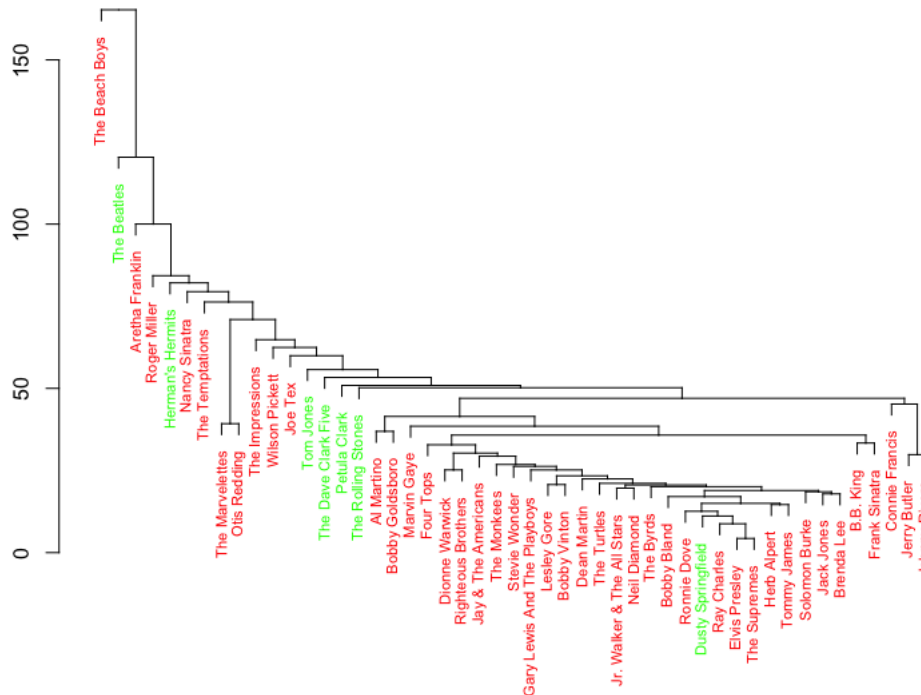**Hierachical Clustering of Lyrical Similairties**

Figure 1: This frog was uploaded via the project menu.

significantly different from their American peers. Interesting findings include the similar tf-idf distribution between the Beatles and the Beach Boys, two of the most successful bands from the 1960s.

Sentiment Analysis

Each song has its associated sentiment. Such sentiment could be labeled to be positive, negative, or neutral. A sentiment analysis is conducted by applying dictionary-based method to compare the distributions of sentiment scores calculated using the lyrics of the song performed by British Invasion artists and their American peers. The dictionary by Jeffrey Breen [1] ,which includes 4783 negative and 2006 positive words is used to calculate the sentiment score. The overall sentiment of a song is positive if its sentiment score[2] is greater than 0. Similarly, if a song's sentiment score is less than 0, it has a negative sentiment. Results are shown as follow:

---

[1]https://github.com/jeffreybreen/twitter-sentiment-analysis-tutorial-201107

[2]sentiment scores = # of positive - # of negative

	British Invasion	American peers	All
Min	-39	-20	-39
Median	7	7	7
Max	45	55	55
Mean	8.24	8.14	8.17
Standard Deviation	11.19	10.94	11.01
Positive (Count)	227	635	862
Postive (%)	81.07	77.25	78.22
Negative (Count)	53	187	240
Negative (%)	18.93	22.75	21.78
Total (Count)	280	822	1102

The distributions of positive/negative lyrics by some of the most representative artists from two groups indicates that, although the average sentiment by the two groups are similar, the scores varies a lot by artists across the two groups.

	The Beatles (BI)	The Rolling Stones (BI)	Dionne Warwick (AM)	The Beach Boys (AM)
Positive (Count)	43	10	7	16
Positive (%)	89.58	37.5	31.82	84.21
Negative (Count)	5	6	15	3
Negative (%)	10.42	62.5	68.18	15.79
Total (Count)	48	16	22	19

Based off the sentiment scores calculated, there is not enough evidence to say that whether the songs performed by British Invasion artists were more positive than that of their American peers.
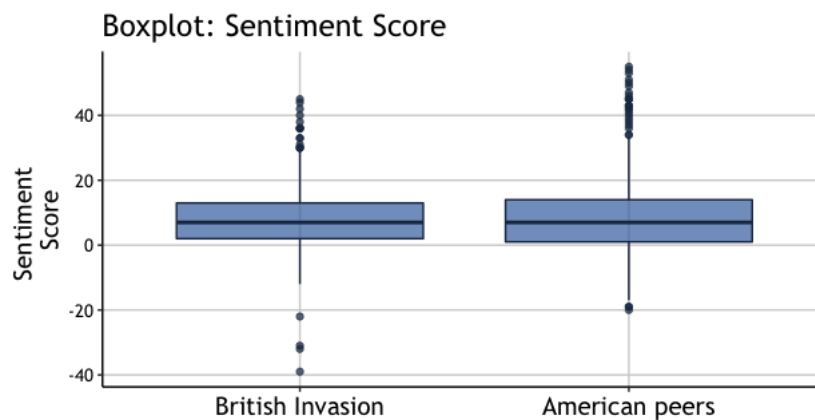

Figure 2: sentiment boxplot

Topic Models

Latent Dirichlet Allocation (LDA) [1], which treats each document as a mixture of a set of topics and each word as attributes to one of the document's topics, is used to extract and interpret the topics of a large text corpus. An alternative method to LDA, Structure Topic Model (STM), is a model that incorporates information about the document to make the results more revealing[5].

Both LDA and STM are conducted in order to detect and compare the main themes of each song. Collectively, the most common theme(s) among all songs performed by a particular artist is considered the most relevant topic(s) that this artist sings about.

LDA

Through the per-topic words distributions B aggregated on the result of LDA C, we identify the topics as:

Topic	Name	Top Words
1	First-person Narrative of Emotions	i, my, know, say, time, yes, want, ever, life, ...
2	Second-person Narrative of Emotions	you, babi, your, know, want, now, like, just, ...
3	Carefree	day, tear, dream, cri, rain, away, summer, ...
4	Good Expectation of Future	go, let, get, round, time, around, hey, everi, ...
5	N/A	come, we, danc, night, bar, back, street, ...
6	Strong Emotion Fueled by Repetitive Words	shake, chain, hurdi, gurdi, amen, funni, ...
7	Positive Experiences	girl, good, got, littl, ain't, sweet, give, ...
8	Love	love, need, world, everybodi, never, arm, ...
9	"He"	he, man, one, him, his, woman, take, can, boy, ...
10	"She"	her, she, look, now, see, mine, just, goin, ...

In topic 1, First-person Narrative of Emotions, is about the singer expressing his/her feelings in a first-person tone. Typical songs that this topic best represents includes "My Generation" by the Who, "I Wang You" by Bob Dylan. In "My Generation", the Who declares a sense of pride for the "generation" that they sings about.

Unlike topic 1, topic 2, Second-person Narrative of Emotions, emphasizes more on how "you" make "me" feel. The lyrics of "I Thank You" by Same & Dave demonstrates this concept: "You didn't have to love me like you did / But you did, but you did/ And I thank you".

Topic 3 embodies a carefree mentality and best represents songs like "Sugar Town" by Nancy Sinatra and "Always Together" by the Dells.

Topic 4 is about the singer's positive expectation for the future, which can be heard from "Tell It On The Mountain" by Peter, Paul & Mary, and the famous "Hey Jude" by the Beatles.

Topic 5 is a mixture of different concept and there is no concise way to summarize this topic. Except "Deadend Street" by the Kinks, emotions and sentiments by the songs that this topic best represents are generally positive.

Songs that are best represented by topic 6 are those with one or several words that are repeated numerous times in the song, such as "Shake" by Otis Redding, in which the word "shake" permeates the entire song, and "Chain of Fools" by Aretha Franklin, where "chain, chain, chain ..." dominates the chorus sections of the song.

Topic 7, Positive Experiences, best represents songs like "Nothing's Too Good for My Baby" by Stevie Wonder.

Topic 8 is predominantly about "love", and is best illustrated by "All You Need is Love" by the Beatles.

Topic 9, "He", best represents songs in which the singer talks about male figures.

Topic 10, "She", best represents songs in which the singer talks about female figures.

Here, "Primary Topic" of a song is defined by the topic that best represents the song. Among all the songs, an artist's "Prevalent Topic" is defined by the majority of the all the Primary Topics of songs this artist performs. In other words, if a song is best represented

by topic 10, its "Primary Topic" is topic 10. Furthermore, if 70% of all the song's Primary Topics among a particular artist's songs is topic 10, then this band's Prevalent Topic is set to be topic 10.

The graphs below shows the distribution of Primary Topics3 and Prevalent Topics4. Compared to songs by the other artists in this corpus, much fewer songs performed by the British artists are best represented by topic 7. On the other hand, more songs performed by the British artists have topic 10 as their Primary Topic. None of the British artists has topic 7 as their Prevalent Topic. The proportion of topic 10 as Prevalent Topic among British artists is higher than that among the other artists.

From the distribution of Prevalent Topics, British artists tend to tell a story in their first-person narrative (topic 1), about love (topic 8), and "her" (topic 10). Perhaps the British culture and the gentle, humble and pleasant image that the early British Invasion bands maintain affected their decisions to not explicitly "show off" or "brag" about a good experience in their works. The best match of a song performed by a British band that can be better represented by topic 7 is "Ain't She Sweet" by the Beatles. However, the song is composed in the United States during the Tin Pan Alley era (1920s).
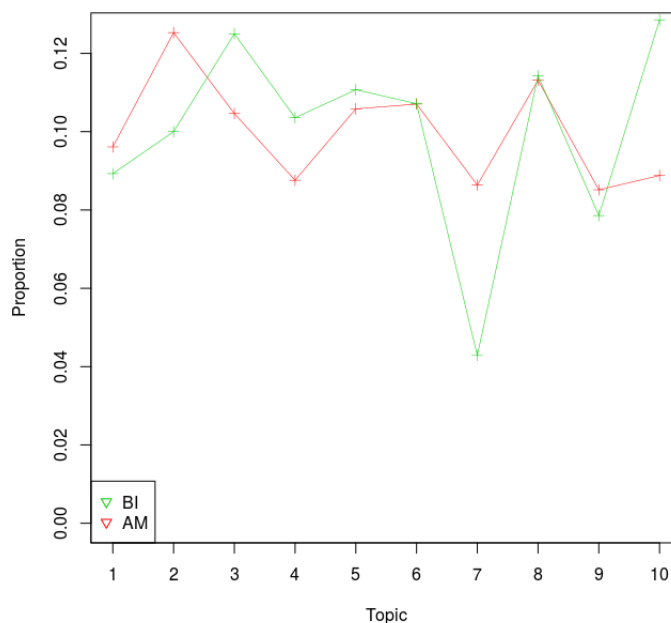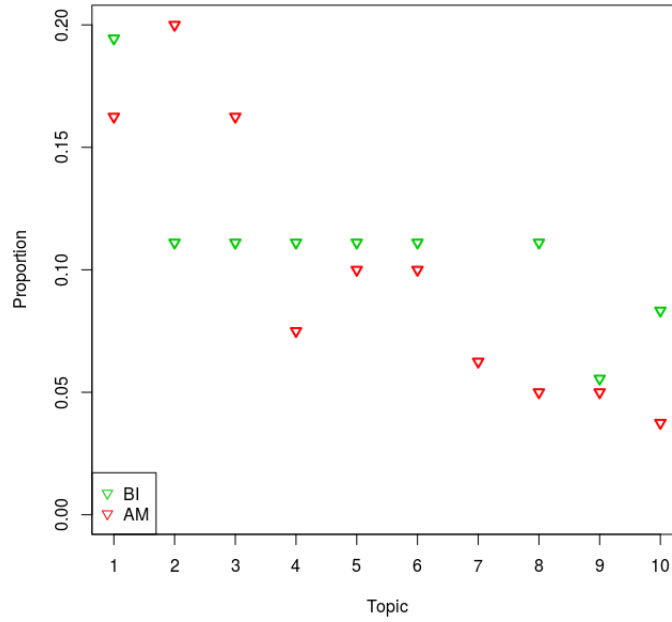

Figure 3: Song-Topic Distribution

Figure 4: Artist-Topic Distribution

STM

The result by STM is less interpretable than LDA[3].

Conclusion and Discussion

Summary

The lyrics of songs performed by British artists whose works have frequently been on the Billboard Hot 100 weekly charts from 1964 to 1969 are not significantly different from the lyrics of songs performed by non-British artists. The result suggests that, lyrics by the Beatles might be more similar to the lyrics by Beach Boys, an American rock band, than to that of other British artists active during the British Invasion Era.

The score distribution from the result of sentiment analysis on the lyrics by British artists and non-British artists suggests that the sentiments that the songs performed by this two group of artists are not significantly different from each other.

Result of topic modeling using Latent Dirichlet Allocation (LDA) indicates that British artists during the British Invasion era tend not to explicitly narrate a very positive ex-

---

[3]Detailed information is attached in our project repository.

perience. Instead, compared to their American counterparts, songs performed by British artists tend to talk more about "Love" and "her".

Limitation and Future Works

There might be other tools to better measure and visualize similarity and difference between lyrics. It is common to have a large portion of repeated words or phrases in lyrics, and although repetitions do have impacts on the importance of the words that are repeated, we believe that the magnitude of such impact must diminish beyond a certain threshold. To some extent, repetitive words affect the accuracy of document-term matrix. Songs have disparate forms from our daily conversations in spoken or written forms. In the future, research efforts can be directed to learning the balladlike structures (Verse-Chorus-Verse-Chorus, Verse-Verse-Chorus-Verse, etc.) and how different words in different positions and song structures can be given adjusted weights in order to better reflect their importance to the document.

Appendix

A Artist Impact

Artist	Category	Impact
The Beatles	BI	56
Dionne Warwick	AM	29
The Beach Boys	AM	27
The Impressions	AM	25
Aretha Franklin	AM	24
The Dave Clark Five	BI	24
Joe Tex	AM	24
Otis Redding	AM	24
The Rolling Stones	BI	23
Al Martino	AM	23
Bobby Vinton	AM	23
Wilson Pickett	AM	23
The Temptations	AM	22
Ray Charles	AM	22
Four Tops	AM	21
Dean Martin	AM	20
Ronnie Dove	AM	20
Frank Sinatra	AM	19
Petula Clark	BI	19
Johnny Rivers	AM	19
Herman's Hermits	BI	19
Bobby Goldsboro	AM	18
Tom Jones	BI	18
Tommy James And The Shondells	AM	17
Herb Alpert & The Tijuana Brass	AM	17
Brenda Lee	AM	17
Jack Jones	AM	17
Marvin Gaye	AM	17
Dusty Springfield	BI	17
The Monkees	AM	17
Lesley Gore	AM	17
Nancy Sinatra	AM	16
B.B. King	AM	16
Elvis Presley	AM	16
Jerry Butler	AM	16
Bobby Bland	AM	16
The Turtles	AM	16
Stevie Wonder	AM	16
The Righteous Brothers	AM	16

B   Top15 Words For Each Topic

Figure 5: Top 15 Words For Each Topic

C   Top5 Most Related Songs For Each Topic

Artist	Song	Category	Topic
The Beatles	Hello Goodbye	BI	1
Jerry Butler	I Dig You Baby	AM	1
Johnny Cash	A Boy Named Sue	AM	1
Bob Dylan	I Want You	AM	1
The Who	My Generation	BI	1
Sam & Dave	I Thank You	AM	2
The Turtles	You Baby	AM	2
Bobby Goldsboro	I Know You Better Than That	AM	2
The Kinks	You Really Got Me	BI	2
The Temptations	The Way You Do The Things You Do	AM	2
Chad & Jeremy	Yesterday's Gone	BI	3
Stevie Wonder	Shoo-Be-Doo-Be-Doo-Da-Day	AM	3
Nancy Sinatra	Sugar Town	AM	3
The Dells	Always Together	AM	3
Ronnie Dove	Happy Summer Days	AM	3
The Temptations	Don't Let The Joneses Get You Down	AM	4
Peter, Paul & Mary	Tell It On The Mountain	AM	4
The Beach Boys	I Get Around	AM	4
The Beatles	Hey Jude	BI	4
Petula Clark	Round Every Corner	BI	4
Roy Orbison	Let The Good Times Roll	AM	5
The Kinks	Deadend Street	BI	5
The Beatles	Yellow Submarine	BI	5
The Beach Boys	Barbara Ann	AM	5
The Beach Boys	Dance, Dance, Dance	AM	5

Artist	Song	Category	Topic
Donovan	Hurdy Gurdy Man	BI	6
The Swinging Blue Jeans	Hippy Hippy Shake	BI	6
The Impressions	Amen	AM	6
Otis Redding	Shake	AM	6
Aretha Franklin	Chain Of Fools	AM	6
Stevie Wonder	Nothing's Too Good For My Baby	AM	7
The Beach Boys	Good Vibrations	AM	7
The 5th Dimension	Sweet Blindness	AM	7
Sam Cooke	Good News	AM	7
Jan & Dean	The New Girl In School	AM	7
The Association	Everything That Touches You	AM	8
The Beatles	All You Need Is Love	BI	8
Nancy Sinatra	Love Eyes	AM	8
Gladys Knight And The Pips	Everybody Needs Love	AM	8
The Mamas & The Papas	For The Love Of Ivy	AM	8
The Kinks	Dedicated Follower Of Fashion	BI	9
Dusty Springfield	Son-Of-A Preacher Man	BI	9
The Kinks	A Well Respected Man	BI	9
Joe Tex	A Sweet Woman Like You	AM	9
Roger Miller	Vance	AM	9
Petula Clark	Look At Mine	BI	10
Jan & Dean	Drag City	AM	10
The Beach Boys	Fun, Fun, Fun	AM	10
Manfred Mann	Do Wah Diddy Diddy	BI	10
Gary Lewis And The Playboys	She's Just My Style	AM	10

References

[1] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[2] Gary Burns. Trends in lyrics in the annual top twenty songs in the united states, 1963–1972. *Popular Music & Society*, 9(1):25–39, 1983.

[3] Charlie Gillett. *The sound of the city: The rise of rock & roll*. Souvenir Press, 2011.

[4] Patricia Fox Ransom. Message in the music: Do lyrics influence well-being? 2015.

[5] Margaret E Roberts, Brandon M Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G Rand. Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58(4):1064–1082, 2014.