– Independent Work Report Fall 2019 –

Tracing Sentiment Change for Careers in Music

Ezra Zinberg
Adviser: Christiane D. Fellbaum

December 18, 2019

Abstract

Sentiment analysis methods allow us to make quantitative claims about sentiment in diverse bodies of text, including music lyrics. We employ term sentiment scores designated by Sentiwordnet to trace the change in negative sentiment in songs by 485 musical artists and evaluate whether negative sentiment increases over the course of the average musical career. Our computations show that negative sentiment increases on average by 3.28% from songs in artists' early career to songs in their late career. While sentiment is demonstrated to change in the hypothesized direction, we found these results to be inconclusive due to the limited magnitude of the sentiment change. We visualize the sentiment change using the Sentiment Triangle plot for 3-coordinate sentiment values.

1. Introduction

1.1. Sentiment Analysis

With the advent of the era of big data and specifically the ocean of human language that is typed and uploaded to the internet every day has come the need for tools to process and make sense of this data. Data scientists, programmers, and linguists have joined forces to develop new structures and systems for analyzing natural language.

One area of natural language processing that has been subject to a good deal of study is the field of sentiment analysis. Sentiment analysis refers to the systematic study of natural language to gain insight into the opinion or emotion of the author or speaker. Individual words may evoke a certain feeling when read or heard, and words or phrases in the context of other terms may indicate the author's opinion of that term. Some words are more sentimentally 'charged', while others may not carry much sentiment at all. In the following sections, we will discuss how particular researchers have classified and organized the various characteristics of a word's sentiment.

Known also as opinion mining, this discipline has been employed in numerous applications, from customer service, to social media monitoring, to bias detection in news articles. Aside from the commercial applications, sentiment analysis has proven useful in in the academic world for processing language found in literature and the humanities at large.

1.2. The Dynamic Musical Career

Given that the tools of sentiment analysis can help us make sense of and find patterns in bodies of text, the study of music lyrics would seem to be a natural application of these methods. Indeed, lyrics as corpora would presumably contain particularly sentimentally charged words, when compared to ordinary, everyday prose. Intuition would state that, on the whole, musicians likely sing about topics of strong emotional quality–both positive and negative emotions–over more mundane subjects. Song is a common outlet for humans to express and release emotion, as much for professional vocalists as for amateurs singing in the shower.

Aside from the easily apparent fact that song lyrics employ sentimentally charged words, we can also observe that the subjects a particular musician sings about can change significantly over time. All around the world of professional music, writers and performers go through phases of their musical careers; perhaps their early songs and late songs can be characterized by different topics, moods, or themes. One must look no further than some of the most famous and influential pop music artists of the recent past, such as Taylor Swift or Kanye West, to notice distinct styles and themes during different periods of the artist's career.

1.3. Goal

The methods developed in the field of sentiment analysis grant the means to investigate some of these intuitions systematically and quantitatively. Studying the calculable sentiment in music lyrics allows us to make more precise and grounded statements about sentiment trends in the world of music.

This paper describes the project of using modern sentiment analysis tools to determine how the sentiment in a musician's songs change over the course of the musician's career. To answer this question and discover general trends in the industry, we processed lyrics from 20,764 songs by 485 different artists using Sentiwordnet to calculate the change in lyrical sentiment over the course of individual artists' careers. Ultimately, we found the average change in sentiment for all artists in the corpus.

2. Related Work

2.1. Opinion Mining

One early effort to make quantitative sense of human emotions and their relationships to one another was initiated by James Russell in his paper, "A Circumplex Model of Affect". In it, the author presents evidence that, rather than different emotions occurring in differing degrees independently of one another, they are actually related in way that can be modeled by the 360 degrees in a circle.[1] [1] In other words, data gathered from subjects showed that 28 different "affects" could be plotted on a two dimensional plane, where the horizontal axis represents degrees of pleasure and displeasure, and the vertical axis represents degrees of arousal and sleepiness.

This work set a precedent for a useful coordinate system in the field of sentiment analysis. It will go on to be adapted by later researchers aiming to quantify sentiment.

Indeed, one of the most significant developments in the area of opinion mining was the construction of Sentiwordnet by Andrea Esuli and Fabrizio Sebastiani. The project gave sentiment scores to all words found in WordNet, the lexical database used widely in computational linguistics.[2] The authors scored these words' sentiment values on a three coordinate system; each word gets a score representing its positive sentiment, negative sentiment, and objectivity.[3] [2] This scheme is based on the idea of two distinct axes for describing a term's sentiment: subjective-objective polarity and positive-negative polarity. The more sentimentally charged or opinionated a term is, the higher it is

[1] Russell, p. 1161
[2] Wordnet. https://wordnet.princeton.edu/
[3] Esuli and Sebastiani, p. 418

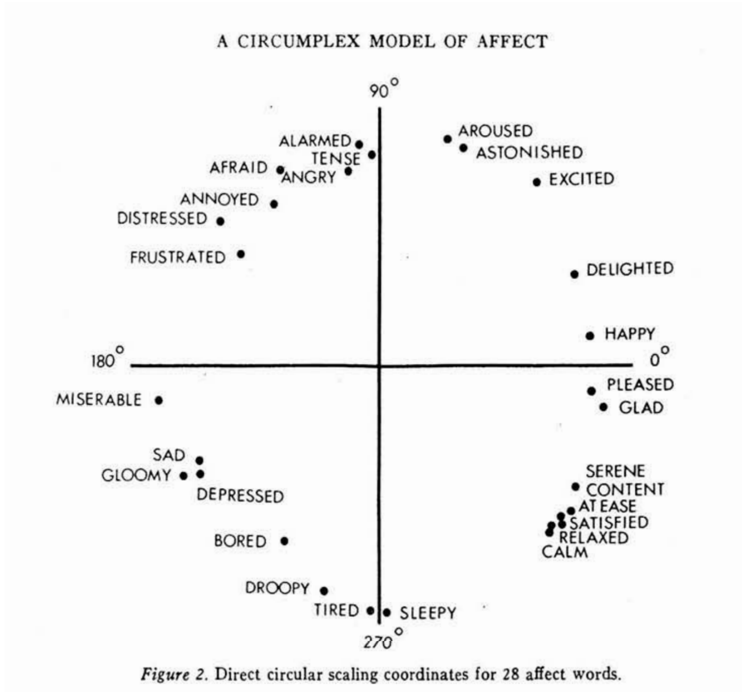

Figure 1: Figure from Russell (1980) showing the positions of affects in the circumplex model

placed on the SO-polarity axis, and vice versa. Each of the three coordinates for a word have take values from 0 to 1 with the property that the objectivity score is 1 minus the sum of the positive and negative scores:

$$(pos, neg, obj) = (pos, neg, 1 - [pos + neg])$$

$$pos + neg + obj = 1$$

This approach is sensible given that the more positive or negative sentiment a word has, the more subjective, and therefore, the less objective it will be.

Word	Positive	Negative	Objective
“happy”	0.875	0.0	0.125
“sad”	0.125	0.75	0.125
“table”	0.0	0.0	1.0

Table 1: Examples of entries in Sentiwordnet with corresponding sentiment scores.

Designating three values for each word in the database leads to the interesting property that every entry in Sentiwordnet can be plotted on a triangular graph, with one axis for each type of sentiment score. A depiction of the "Sentiment Triangle" was provided in the 2006 paper describing Sentiwordnet. (See Figure 2)

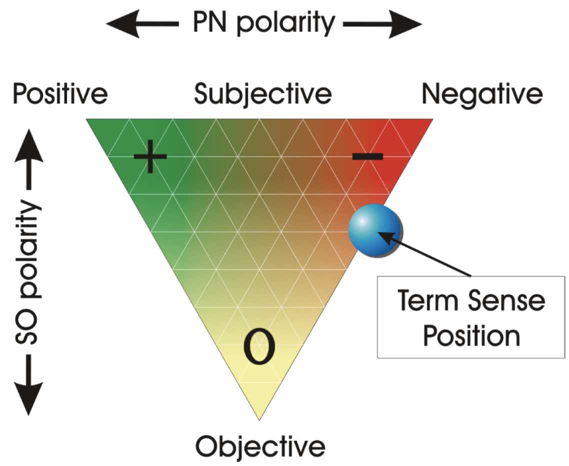

Figure 2: Sentiment Triangle from Esuli and Sebastiani (2006)

2.2. Lyrics Analysis

Researchers in numerous fields have examined and reasoned about song lyrics, and as the discipline of natural language processing has matured, have used some of the methods from that discipline in their work.

An early analysis of lyrics in pop music was performed by Tim Murphey, described in an article from 1992, entitled "The Discourse of Pop Songs". [3] Looking at a corpus of 50 top pop songs from 1987, the author made observations about songs' words-per-minute and the songs' use of time, place, and gender, among other factors. This work is an early example of quantitative analysis of music lyrics. While it studied a small corpus and analyzed a limited number of factors, the paper showed the potential for interesting conclusions to be made from studying patterns in lyrics.

More recently, researchers at the Polytechnic University of Turin embarked on a project to create a data base of songs that would be tagged by one of four moods: happy, angry, sad, and relaxed. [4] They drew from a combination of lexica–WordNet and others–to process lyrics from about 2,600 songs and constructed their mood-tagged lyrics database, MoodyLyrics. Most significantly, the authors concluded that when they asked human subjects to tag songs with one of the four moods, the human tagging agreed with the computer-generating tagging in 74.25 percent of cases. This figure of accuracy with respect to the 'gold standard' of human surveys confirmed the reliability of computational methods in applying sentiment analysis to music lyrics. We can be reasonably confident that sentiment analysis methods produce results that align with the sentiment that humans perceive when listening to music.

3. Approach

3.1. Motivation

The application of sentiment analysis to a corpus of music lyrics enables us to investigate intriguing questions in the world of music, and perhaps even human psychology more broadly. The strength of a computational approach to studying lyrics specifically and natural language in generally is that programs provide real, quantitative insights about huge amounts of text, which before we may only have been able to describe through surface-level observations, intuitions, and anecdotal evidence.

In particular, we can address the question of how a musical artist's style changes in terms of sentiment from the beginning of their career until its end. The progression of a musician's lyrical content over the course of their career and they way it makes listeners feel seems to be a phenomenon that pop culture acknowledges and accepts without data-backed evidence. Examining this phenomenon through the lens of sentiment analysis has the potential not only to quantify existing intuitions about musicians, but also to uncover new insights into trends in music.

3.2. New Questions, New Data

Our approach lies at the intersection of several developments in natural language processing and builds upon that work. However, we address questions that have not been addressed by previous research–namely, how does sentiment change for musical artists from their early work to their late work? While "The Discourse of Pop Songs" was an early example of an academic analysis of pop lyrics, it worked with a small corpus of only 50 songs and did not apply any opinion mining methods. "MoodyLyrics", a more recent effort to apply sentiment analysis to music lyrics, worked with a somewhat larger corpus of 2,600 songs, but did not focus on individual artists as a object of study.

Our work also accounts for a much larger corpus of music lyrics than either of the aforementioned projects. Both in terms of songs (20764) and artists (485), our research represents a significant sample size of the data and therefore provides more comprehensive results.

On the whole, we focus on a unique combination of questions and data sets that may grant new insight into the field of sentiment analysis in music.

3.3. Hypothesis

We hypothesize that, on average, over all artists, negative sentiment in an artist's lyrics will increase as the artist's career progresses. Below (section 4.2), we describe the evaluative metrics we use to quantify sentiment change.

The null hypothesis and alternative hypothesis are defined as follows:

H_0 : *On average, there is no significant increase in negative sentiment from an artist's early work to their later work.*

H_1 : *On average, there is a significant increase in negative sentiment from an artist's early work to their later work.*

The coming sections outline the process by which we tested this hypothesis.

4. Methods

4.1. Strategy

The procedure used to trace change in sentiment for a sizable number artist's entailed several high-level steps. First, we acquired a large enough sample size of songs by a large enough variety of musicians to serve as the object of the analysis. After preprocessing this data set and preparing it for testing, we calculated a number of figures for each artist in the corpus. To measure change in sentiment from an artist's early lyrics to their later lyrics, we produced values for each artist's

(1) overall sentiment average, (2) sentiment average in their early work, and (3) sentiment average in their late work. Once we calculated the necessary figures for each artist, we applied them to our evaluation metrics and generated sentiment change values for each artist. Finally, we took an average of the individual artist metrics over the 485 total artists to present figures representing the music industry as a whole.

4.2. Evaluation Scheme

We evaluated whether our data supported or contradicted our hypothesis by constructing a set of three metrics that describe an artist's lyrical sentiment change. Each metric expresses a different type of sentiment change, each providing a measure of sentient change for different portions of the musician's career. We define:

$$\Delta S_{total} = S_{late} - S_{early}$$

$$\Delta S_{early} = S_{avg} - S_{early}$$

$$\Delta S_{late} = S_{late} - S_{avg}$$

where S represents the negative sentiment score, ΔS represents change in negative sentiment, S_{avg} represents overall average negative sentiment, S_{early} represents sentiment in the early period, and S_{late} represents sentiment in the late period.

For each of the three metrics, the greater the value, the more negative sentiment increased from an earlier point in the artist's career to a later point. When averaged over all artists, we assess the validity of the hypothesis by stating that if the ΔS values are positive, H_1 is supported. If the ΔS values that emerge are negative or not significantly positive, then H_0 is supported.

4.3. Corpora

The lyrics analyzed for this research come primarily from a single large database of modern song lyrics, supplemented with additional data fields from a secondary database.

Our primary data set is the Song Lyrics database provided by Kaggle, originally sourced from

LyricsFreak.[4] [5] [5] The full data set contains fields for artist, song title, and song lyric for 57,650 songs by 643 unique artists. However, to test our hypothesis, we also needed a field for song release date so that we could categorize a song as an artist's early work or late work appropriately. To acquire release date data, we merged the Kaggle data set with data collected from the MusicBrainz-Python API.[6] [6] After matching the songs in the Song Lyrics database with entries from the MusicBrainz database, the total number of usable songs decreased to 20764 and the number of artists fell to 485, because the rest of the songs present in the original data set had no release date data in the MusicBrainz data set. Ultimately, our results were calculated based on the smaller, unified corpus of approximately 20,000 songs.

Metallica	Stone Dead Forever	8/18/08	And didn't you hear me you never listen past And didn't I see you gone to seed And the only reason is that you're too young greed
Metallica	Wherever I May Roam	6/28/06	I have stripped of all but pride So in her I do confide And she keeps me satisfied
Metallica	Whiskey In The Jar	2014-07	The Cork and Kerry Mountains I saw Captain Farrell And his money, he was countin'
Michael Bolton	A Heart Can Only Be So Strong	1997	Just to walk back a hundred times more A thousand nights, I have sworn not to stay, ooh oh I remained tangles in the chains of desire, ooh
Michael Bolton	A Love So Beautiful	1995	Our love long ago But in my heart I feel the same Old afterglow
Michael Bolton	A Time For Letting Go	2005	Isn't really what you need And the dream and all its promise Was never meant to be

Figure 3: A sample of entries in the unified data set, with fields for artist, title, release date, and lyric string.

4.4. Implementation

After downloading the song lyrics data set, we wrote a Python script that takes a song lyric string as input and returns values for the song's average positive, negative, and objective sentiment. Within this program, the lyric data is cleaned by converting the words in the lyric to tokens that can be matched with words in Sentiwordnet and removing stop words that clutter the more significant words in the song. This preprocessing is executed by function calls to the Python Natural Language

[4] Kaggle Song Lyrics database. https://www.kaggle.com/mousehead/songlyrics
[5] LyricsFreak. https://www.lyricsfreak.com/
[6] MusicBrainz.org. https://musicbrainz.org/

Toolkit library.[7] Each token in the song is then passed as an argument to the Sentiwordnet library included in the NLTK API.[8] When given a word, Sentiwordnet returns senitment scores for all definitions of the word found in WordNet. To circumvent the issue of determining which definition is being used in the context of the song, which would have required additional layers of complexity, we decided to consistently utilize the first returned definition, which is guaranteed to be the most common usage of the word. This allows us to come up with average sentiment values for all tokens in the song lyric.

The described Python function is, in turn, called by another Python program that iterates through the CSV format file containing the lyrics data and performs this sentiment analysis on all songs by a given artist. From the artist's overall sentiment data, we calculate S_{avg}, which is needed to find ΔS_{early} and ΔS_{late}. To calculate the three evaluative metrics, we designated an artist's early work to be all songs released in the first fifth of their career (i.e., if an artist's first release was in 2000, and their last release was in 2010, then the first fifth of their career is the period 2000-2002). Likewise, late work was designated as songs from the last fifth of the artist's career. We could then generate figures for S_{early} and S_{late}. Having found S_{avg}, S_{early}, and S_{late}, we calculated the three main evaluative metrics for each artist.

An additional Python script allows us to call the above function and repeat the process for every artist in the database. We constructed a CSV file enumerating all 485 artists along with values for their three sentiment metrics. Iterating through this list, we calculated the average of these sentiment metrics over all artists.

The implementation was executed with the help of several intermediate Python programs and CSV files.[9] At some stages, acceptably small fractions of the data had to be disregarded due to lack of supporting information for those particular entries–e.g., songs that were present in the Kaggle database but did not have a release date in the MusicBrainz database. However, the final data set contained over 20,000 entries and was judged to be a large enough sample size for our research.

[7] Natural Language Toolkit API documentation. https://www.nltk.org/

[8] Sentiwordnet API documentation. https://www.nltk.org/api/nltk.corpus.reader.html

[9] A repository of these supporting files can be accessed at https://github.com/ezinberg/SentiLyric.

5. Results

We arrived at the following results. In this case, percent change is defined at the proportional difference between that value and the average value for negative sentiment over whole careers for all artists. Let this overall average be defined as S_{AVG}. We calculated the % Change field by this formula:

$$\% \; Change = \frac{value}{S_{AVG}} \cdot 100$$

We found that:

$$S_{AVG} = 0.0458$$

The figures for our three evaluative metrics over all artists are presented in the table below.

Metric	Definition	Value	% Change
ΔS_{total}	$\Delta S_{total} = S_{late} - S_{early}$	0.0015	+3.28%
ΔS_{early}	$\Delta S_{early} = S_{avg} - S_{early}$	0.0011	+2.40%
ΔS_{late}	$\Delta S_{late} = S_{late} - S_{avg}$	0.0004	+0.88%

Table 2: Experimental results. Average sentiment change values for all artists in the database.

6. Discussion

6.1. Reasoning About the Results

The above results support the hypothesis, but only to a small degree. As evident from the % Change figures, negative sentiment increased–changed in the positive direction–for all three evaluative metrics. As all three metrics denote negative sentiment in an earlier point in the career subtracted from negative sentiment in a later point in the career, positive values indicate higher negative sentiment toward the end of the career compared to negative sentiment earlier in the career, which is what our hypothesis predicted. However, despite emerging as greater than zero, each of the values are somewhat small quantities. In other words, the direction of the sentiment change is consistent with the hypothesis, but the magnitude of the sentiment change shows that negative sentiment does not increase by a substantial amount. The relationship between each of our experimental results and our hypothesis is summarized in the table below.

Metric	Value	% Change	Supports H_1?
ΔS_{total}	0.0015	+3.28%	Yes*
ΔS_{early}	0.0011	+2.40%	Yes*
ΔS_{late}	0.0004	+0.88%	Yes*

Table 3: Do the results support H_1?
***Not significantly**

We may also reason about the relationships between the three metrics. Predictably, the sum of ΔS_{early} and ΔS_{late} is virtually the same as ΔS_{total}, as ΔS_{early} and ΔS_{late} simply represent total sentiment change partitioned into the differences between early or late sentiment and average sentiment.

Additionally, we can observe that ΔS_{early} emerged as approximately three times the magnitude of ΔS_{late}. This could indicate that the increase in negative sentiment is generally more substantial in the early portion of a career. Perhaps we can speculate that during the beginning of a career in music, an artist tends to explore and experiment with their content more before they settle in to their lyrical style toward the middle and end of their career. This could explain why the negative sentiment value appears more volatile in the early career.

Overall, these results show weak evidence in favor of the hypothesis.

6.2. Visualizing the Results

We make use of the Sentiment Triangle from the paper describing Sentiwordnet to help visualize our results. On this graph, sentiment values are plotted as points, which can represent words or average values over many words. In the same vein, change in sentiment is shown as a trajectory moving from one sentiment point to another.

For the purposes of this project, we dealt exclusively with negative sentiment, ignoring the other

Sentiwordnet fields for positivity and objectivity. Nevertheless, all three fields are accounted for when visualizing this data on the Sentiment Triangle.

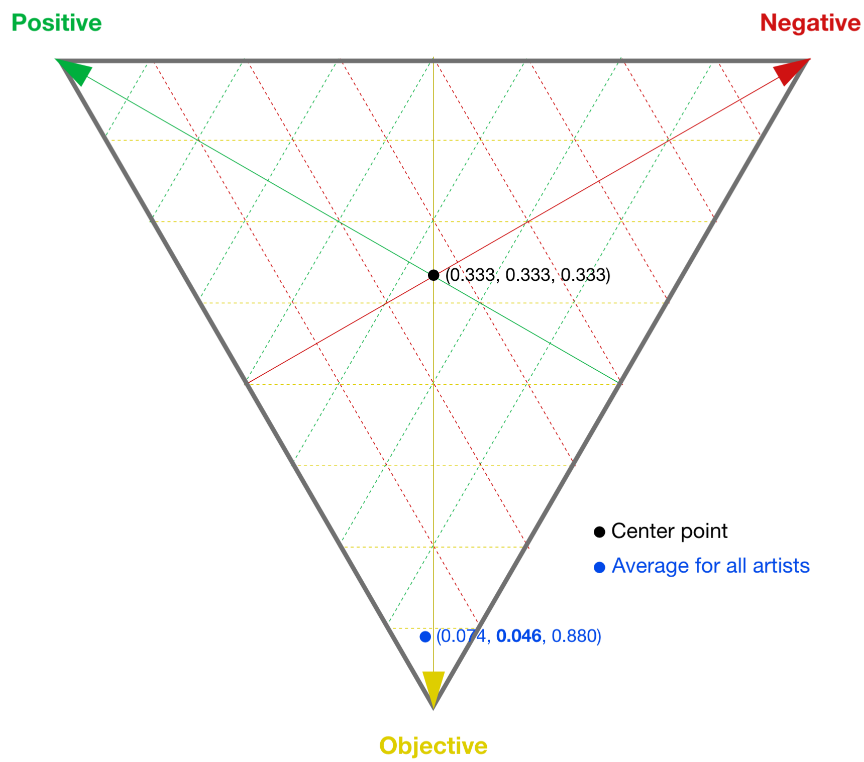

Figure 4: The Sentiment Triangle. Our discussion centers around the overall average sentiment point, which has coordinates (0.074, 0.046, 0.880).

On the scale of the full Sentiment Triangle, the sentiment change observed in music lyrics over time is very subtle. As apparent from the location of the average point, the vast majority of words in the lyrics we analyzed skew toward the "Objective" vertex of the triangle. (See Figure 4)[10] This is likely because a significant number of words in WordNet, and in the English language, are not laden with much sentiment, but are simply more objective words. Still, this outcome is surprising, given that we might have assumed music lyrics would be particularly sentimentally charged, compared to a more generic corpus of prose, for example.

To better visualize the sentiment change we found in this research, we can take a closer look at the relevant area of the triangle.

[10]The location of points on the graph are approximate.

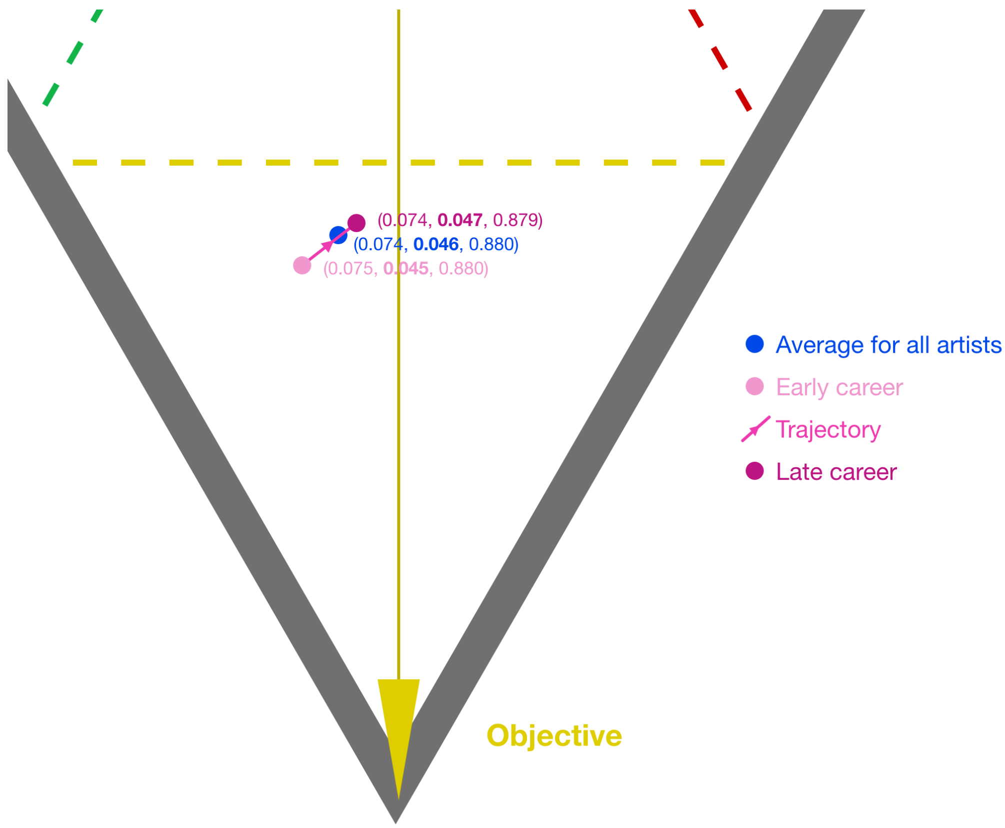

Figure 5: The bottom portion of the Sentiment Triangle, showing the approximate trajectory of the average artist's sentiment. Negative sentiment in these coordinates in boldface.

Figure 5 shows the sentiment trajectory of the average artist from the beginning to end of their career. Starting at the bottom left sentiment point, sentiment in the early career, the artist's lyrical sentiment tends to travel along the shown trajectory, ending at the top right point, lyrical sentiment in the late career. From the early career to the late career, sentiment gravitates toward the "Negative" vertex of the triangle, and slightly away from the "Positive" and "Objective" vertices. Zooming in on the bottom area of the triangle helps illuminate both the direction of the trend toward more negative sentiment over time, and also the limited magnitude with which sentiment changes. The change in these coordinates only becomes apparent on the order of 10^{-4} of a sentiment point, i.e., 0.01% of the total possible sentiment value. We also observe in a visual manner that sentiment is more distant from the average in the early career than in the late career.

Stepping back, the Sentiment Triangle illustrates just how little sentiment changes on average. Sentiment does, indeed, tend to become more negative, but not by much.

7. Summary

7.1. Conclusions

To what extent was our hypothesis confirmed? We have determined that the results of this research are inconclusive in confirming or refuting the hypothesis. Although we showed that sentiment did change in the hypothesized direction, the magnitude of this change was not large enough to confirm that this change was significant. However, these results should not be discounted as meaningless; we still see indications that the intuition that singers address more negative topics as their careers progress is supported by the evidence found in a substantial number of music lyrics.

7.2. Limitations

Several potential sources of error exist in the data sets and sentiment analysis procedure used for this research. First, though our data set of lyrics and artists was fairly large, it cannot perfectly stand in for all lyrics. Second, our use of Sentiwordnet to score tokens from song lyrics assumed that the first definition of that token in Sentiwordnet was the same meaning of the word being used in the context of the song. Other issues emerge when considering the potential discrepancy between the token fed into Sentiwordnet and the actual sentiment intention of the artist. For example, negation of words in context (e.g., “not happy”) would be processed as two separate words, “not” and “happy”. Sarcasm, where the speaker conveys the opposite sentiment of a words literal meaning, was also disregarded in our analysis.[11]

In addition, our definition of early and late career used a coarse-grain view of what constitutes an early and late period in a musical career. (Section 4.4) Musical careers tend to be unpredictable and nonlinear, making it a nontrivial task to create an accurate periodization of such a career, especially so when trying to designate periods automatically for the careers of hundreds of artists.

[11] Though word context was not taken into account for this project, there is an argument to be made that the sentiment in a song felt by the listener is influenced by the literal meaning of the word as well as the intended meaning in context. The difference between the phrases, “I’m sad” and, “I’m not happy” may be that, despite conveying generally the same message, hearing a positively charged word like “happy” could conceivably influence the sentiment of the overall phrase in the positive direction. In other words, while context is key for an ideal sentiment analysis, the literal meanings of words may affect sentiment regardless.

7.3. Future Work

Future sentiment analysis research applied to music lyrics will explore new questions in both fields. Using similar opinion mining techniques combined with data sets like the ones studied in this paper could compare sentiment in music lyrics and sentiment in other types of text, such as prose literature, non-musical poetry, conversational dialogue, etc. The field of sentiment analysis would also benefit from the development of the essential tools of the trade. For example, ensuring that Sentiwordnet is provided the correct definition of a word in context would make the database more reliable and true to the source text. New interfaces and standards for these tools would take better advantage of the data available for natural language processing.

With respect to our particular question, our results could be reinforced by the study of word meaning in context. Such a project could be executed with word embedding technologies or by examination of longer n-grams from the text. There are ample opportunities for further application of sentiment analysis tools to study related topics.

We also acknowledge the difficulty in making claims about musical sentiment in general while disregarding the scores of other factors in a song that affect its sentiment. Lyrical content is only one variable among others like melody, rhythm, musical and lyrical articulation, vocal cadence, and so forth. To support more holistic claims about sentiment in music, inquiries should be done into all the diverse factors that influence sentiment.

References

[1] Russell, James A. "A circumplex model of affect." Journal of personality and social psychology 39.6 (1980): 1161.

[2] Esuli, Andrea, and Fabrizio Sebastiani. "Sentiwordnet: A publicly available lexical resource for opinion mining." LREC. Vol. 6. 2006.

[3] Murphey, Tim. "The discourse of pop songs." Tesol Quarterly 26.4 (1992): 770-774.

[4] Çano, Erion, and Maurizio Morisio. "Moodylyrics: A sentiment annotated lyrics dataset." Proceedings of the 2017 International Conference on Intelligent Systems, Metaheuristics and Swarm Intelligence. ACM, 2017.

[5] Sergey Kuznetsov. "55000+ Song Lyrics: Lyrics for 55000+ songs in English from LyricsFreak". Kaggle. Web. September 18, 2019. https://www.kaggle.com/mousehead/songlyrics.

[6] "MusicbrainzNGS 0.6 Documentation". https://python-musicbrainzngs.readthedocs.io/en/v0.6/.

I pledge my honor that this paper represents my own work in accordance with University regulations.