The Baker Toth Lyrical Creativity Index

William A. Toth and **Charles D. Baker**
Dartmouth College
Hanover, NH 03755
{william.a.toth.23,charles.d.baker.24}@dartmouth.edu

Abstract

One of the most common metrics that music critics use when comparing musicians is their creativity. Artists spend years attempting to push the limits of their own creativity for their listeners. However, musical creativity is incredibly subjective and difficult to analyze. Here we show that we can solve this problem with the Baker-Toth Lyrical Creativity Index. This is a measure of how lyrically (not melodically) creative a song is. The index is a composite score of three different metrics which all analyze lyrical creativity using different algorithms. After a survey of 10 songs we found our creativity index to have an average percent error of 26 percent. This result is quite exciting as it is one of the first indexes of its kind and it is already showing a somewhat reliable trend. These results show that, while difficult, it is still possible to classify the creativity of various songs and artists at an objective level. Our metric, and future lyrical creativity metrics, can also be combined with other creativity metrics that analyze different aspects of songs such as its creative rhythm or melodies. This composite score would represent a more holistic approach to musical creativity than a simple lyrical approach.

1 Introduction

Creativity as defined by Oxford Languages is the use of the imagination or original ideas, especially in the production of an artistic work. It is hard, however, to apply a rigid framework to creativity because of its subjective nature. One of the best ways to express creativity is within music. Artists put their heart and soul into songs, as we often find ourselves wondering who is the most creative. We aim to, with our algorithm, establish a concrete way to determine the creativity level of a given song, artist, genre, decade, and more. By measuring repetitiveness, uniqueness, nonconformity to existing genres, we have created the Baker-Toth lyrical creativity index, a number which represents the total creativity of a given song solely through its words.

There is previous work done in determining what makes a body of text creative, and we wish to extend this to music. Previous work studies the link between characteristics and perceived creativity, noting all of the things that contribute to it in lexical composition (Kuznetsova et al, 2013). What is so interesting about previous study and creativity in general is that the idea is so subjective. For example, when trying to study what makes a work of art creative, we study what makes people think this work is creative, which we draw upon especially when evaluating our program. Previous work also shows that again, because creativity is so subjective, different cultures perceive creativity differently, and any system we create to score it may only align with a certain group's idea of creativity (Myung-Sook, 2013).

We wish to extend this previous work as well as add our own ideas to the music realm. Because this is an NLP algorithm, it does not take into account melody, chord changes, duration, etc. and other musical aspects which measure creativity, but rather is a measure of only lyrical creativity. In the future, we hope to reconcile our measure with more melodic measures to get a holistic score for creativity.

2 Methodology

2.1 The Dataset

The dataset comes from mendeley.com. Each of the 28000 entries is a song including year, genre, lyrics, and scores from 0 to 1 for various topics like danceability, loudness, violence, etc.

2.2 Methodology Overview

To get our composite score, we can average our scores for certain criteria for creativity. We will go more into the math and less the implementation in this section, as that is visible in our code (i.e.

'we want to measure a certain ratio' and not 'we want to create this dictionary and index into it and change it this way'). As a preprocessing step, we remove stop words because those are not indicative of the song's lyrics, and we don't want to punish a song that uses many or reward a song that doesn't use very many.

2.3 Repetitiveness Score

The first element we want to measure is repetitiveness of a given song. A creative song will not repeat itself a lot, and we want to reward songs that are consistently using different words. To do this, we do not need our dataset at all, but rather can look at the specific song. We are interested in something called the type token ratio, whose formula is below

$$\text{TTR} = \frac{\text{uniquewords}}{\text{totalwords}}$$

If the type token ratio is near 1, then we want to reward the song because every word is not repeated. If the ratio is low, then words are very frequently repeated, and we don't want to reward the song. Due to this, the type token ratio can itself be the repetitiveness score for the song.

2.4 Uniqueness Score

The next element for measuring creativity is how unique the words in the song are. This is less easy to measure. What we have done is first calculated the dataset frequency for each unique word in a track. A word that appears a lot in the dataset is a less unique word. A word that seldom appears in the set is a more unique word. We then take the average dataset frequency over all the words in a given song. We want to reward lower frequencies because those correspond to more unique lyrics, so we take the minimum frequency over all the songs and set that score to 1, the maximum and set it to 0, and any number in the middle will end up on a 0-1 scale. The formula for average frequency within the dataset is below.

$$\frac{\sum_{\text{word}\in\text{uniquewords}} \text{datasetfrequency}(\text{word})}{\text{length}(\text{uniquewords})}$$

2.5 Song Cluster non conformity

Finally, we want to measure how unlike other songs a song is. We can think of this as genre nonconformity, although what we are comparing is not genre, but rather groups of other songs. Luckily, our dataset has ranked each song on a scale of 0 to 1 in many categories like danceability, loudness, and more, so what we can do is cluster all of our tracks with these categories as features. We chose 10 centroids, but this is an arbitrary number. What this does is put each song in a centroid, where each centroid are songs that are similar. We can think of each of these centroids as a genre, although they are not exactly that. For a given track, we can then take the distance from each centroid and take the variance of these distances. Mary Lou Maher notes that novelty can be measured as distance from previous artifacts or works (Maher, 2010), and that is exactly what we are trying to measure here.

The reason we take the variance is it measures the nonconformity to centroids. If a track has low variance, that means it's not particularly close to any particular centroid but rather is at a far distance from all of them, and thus isn't like any song group. If it has a high variance, that means it is very close to one or more centroids and it is very far from certain centroids. If this is the case, then it is very like other songs and conforms more to that group. We want to punish high variances and reward low variances, so we assign 0 to the highest variance, 1 to the lowest, and some value 0-1 for those in between based on how close they are to the low variance. The formula for variance of distance from centroids is below.

$$\text{var}(\text{song}) = \frac{\sum_{\text{dist}\in\text{centroiddists}} \text{dist}^2}{\text{length}(\text{centroiddists})}$$

3 Results

3.1 Overview

As for results, there is no objective creativity score for a song to compare to, so our results come in two forms. First, we have the trends and insights we took from our program. Second, We have the performance of our program, in other words the way it did versus human response to these songs.

3.2 Trends and Insights

First, we took an average of the score of each song for any artists that had more than 3 songs. Our most creative included Steps Ahead, wu-tang clan, Lee Morgan, and ll cool j. The least creative included will.i.am, Natasha Beddingfield, Selena Gomez, Lady Gaga, Britney Spears. This was interesting seeing Jazz and rap artists among the most lyrically creative and all pop artists among

the least creative.

To test this further, we averaged the creativities for different genres, and the order of least creative to most creative was pop, blues, reggae, rock, jazz, country.

3.3 Performance

Because creativity is subjective, there is no objective metric to compare it to. To test performance, we decided to test our numbers against the consensus of the people. To do so, we sent out a google form asking people to rank songs' lyrical creativity. We then compared their scores to ours, with the results below.

Song	score	response	%err
Queen of Califor...	0.86	0.74	17
Hoes Mad	0.48	0.38	24
Never Be Like you	0.45	0.53	61
Badlands	0.53	0.68	21
With or Without You	0.73	0.50	44
Mr. Saxobeat	0.47	0.35	34
In My life	0.80	0.67	19
Boulevard of Bro...	0.51	0.61	16
Heartache Med...	0.52	0.55	6
Trap Hop	0.55	0.65	15
Average	N/A	N/A	25.9

4 Discussion

As for trends, sure enough pop is at the bottom, but instead of seeing jazz or hip hop at the top, we see country. This may be a surprise, but if we give it some thought, it makes more sense. We think of country songs as uncreative because they are melodically uncreative, not because they are lyrically uncreative. When we actually look at country lyrics, they are often recounting some story or talking about some anecdote instead of repeating one thing over and over. Also, we have jazz as a close second. Rap gets grouped in with hip hop, which may hurt it's score, but we'd like to see if rap had its own genre.

For performance, our average percent error was around 26%. It is hard to determine how great this number is, but when we think about it, we are trying to use a computer to model the human idea of creativity, which is one of the most complex assessments one can give. Looking at the differences in our scores and the response scores, we do see a trend of our scores being low when those are low, high when those are high. Taking this into context, 26% is actually pretty good. In addition, we simply sent this out to friends and likely didn't have enough respondents for outliers not to affect the average scores. Also, we mentioned to respondents that they should only use lyrics, but a human has a hard time separating lyrics from melody, so our human responses may be flawed and not the best thing to compare to for lyrical creativity. An alternative would be to only use songs poeple hadn[t heard before, so their perception lies in the lyrics. In all, we are happy with the way this performed, given that we are trying to model the human assessment of creativity, something far more complex than what a small NLP project should be able to accomplish.

Ethically, our use of data poses no problems. All of the data is readily available on the internet and the songs are published, so there is no problem with that. However, we don't know how the makers of the dataset scored the themes like violence, danceability, loudness, etc., and their practices for this could have been biased in some way. We also need to acknowledge that this is simply our way of rating creativity, and some other cultures may frame creativity differently. As of now, this isn't a problem, but if our project were to ever be used for professional purposes, it might have bias against certain songs or conceptions of creativity.

5 Conclusion

Broadly, we attempted to model the human assessment of creativity with NLP. Our score for lyrical creativity took into account 3 criteria that we deemed important for a song, and we believe it performed well with an average percent error of 25.9%. It also allowed us to rank artists and genre in a way that was consistent with our consensus (for example, jazz and rock scored higher than pop). As for future work, it is important to know that this creativity score is only for lyrics. It is not a holistic view of creativity, and is rather a small piece in the puzzle. We'd love to see our work reconciled with some sort of algorithm that scores melodic creativity and be used by musicians, magazines, and the everyday person to understand how creative a song is.

6 References

Maher, Mary. (2010). Evaluating creativity in humans, computers, and collectively intelligent systems. DESIRE'10: Creativity and Innovation in Design. 22-28.

Polina Kuznetsova, Jianfu Chen, and Yejin Choi. 2013. Understanding and Quantifying Creativity in Lexical Composition. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 1246–1258, Seattle, Washington, USA. Association for Computational Linguistics.

Auh, Myung-sook. "Assessing Creativity in Composing Music: Product-Process-Person-Environment Approaches." (2013).