Comparing Skip-gram Embeddings Across Pop and Rap Music to Explore the Semantic Variation in Music Genres

John Speaks
jspeaks2@illinois.edu

Abstract

This paper presents a semantic analysis of the lyrics to pop and rap songs. For this, a skip-gram model is trained to create embedding spaces for each genre based on a corpus of genre-labeled data from Genius via Kaggle[1]. The embedding spaced were then compared and explored through the analysis of the neighborhoods of randomly selected words from the rap and pop embedding spaces. In the process of training the embedding spaces, difficulties due to language variation within songs, strange formatting, and informal representations of words occurred, but valuable findings were still possible. The findings of this paper provide interesting, repeatable, and quantifiable insights into the semantic differences between the two genres of music using Jaccard similarity. The entire pipeline used in the paper is available on GitHub[2] for further tinkering and research. This exploration opens the door to more research using the same data but with more resources, better pre-processing, and/or a different focus outside of just the semantic domain.

1 Introduction

Music is a uniquely human experience that varies vastly from how we regularly use language to communicate, and also from other forms of music. With more and more music online, there is also much more ready access to digital transcripts of the lyrics of popular songs. One such corpus is from Genius[3], a website containing the lyrics to many songs and used by its users to learn more about the songs they love.

This paper explores semantic aspects of language used in pop and rap music lyrics collected from Genius. The corpus contains millions of songs across countless genres, but in this paper, we focus on two: pop and rap. These were chosen because they are the largest categories in the Genius data, and because Genius primarily focuses on rap and pop, making them the most reliable categories.

To compare the semantic differences in each music domain, an embedding representation is created for both genres using a Skip-gram model. These embedding spaces are then compared to gain a better understanding of each genre. In the study, many issues were encountered with the formatting of the data and the inherent formatting of music lyrics that do not necessarily align with the direct semantic meaning of the words being represented. Still, valuable analysis was possible from the data.

This paper will follow an organized approach. Firstly, it provides background information on the corpus utilized for analysis, detailing the sampling procedure and pre-processing steps taken to prepare the data for training the embedding spaces. Next, the methodology sections details the application of the skip-gram model to train the embedding spaces. After this, the results of the embeddings are presented and evaluated against each other, providing insights into the semantic patterns of pop and rap music. Finally, new avenues for research are highlighted based on the work done in this paper.

2 Background on the Corpus

The corpus used for this experiment was data collected from the website Genius. Genius has a large collection of song lyrics from a variety of music genres. The dataset was obtained from Kaggle[4], and contains over three million songs and for each, the language that it is in and the full lyrics, a tag for the genre, and labels for the song language among

[1] https://www.kaggle.com/datasets/carlosgdcj/genius-song-lyrics-with-language-information

[2] https://github.com/JTSIV1/Pop-and-Rap-Music-Embedding-Space-Comparison/tree/main

[3] https://genius.com/

[4] https://www.kaggle.com/datasets/carlosgdcj/genius-song-lyrics-with-language-information

other data. The lyrics require a lot of preprocessing because they are in the format that would be displayed on the Genius website. This means that the lyrics contain metadata in line between square brackets, many new line characters, and extra blank lines. Another difficulty that the data provided was that many songs feature several languages. The data was labeled with different languages using FastText's langid algorithm and the CLD3 algorithm. This provides useful data but does not account for the many songs that contain lyrics in more than one language. This is an issue because the aim of this experiment was to compare English embeddings across different music genres, but using the language labels to filter out non-English languages still leads to some other languages in the data.

Another issue faced with the corpus was stylistic misspelling and onomatopoeia in the data. To represent artistic variation in what is being vocalized in the songs, the lyrics contain strings like 'heeeeeeeeeeey' in addition to the word 'hey' which is what is actually being represented in terms of semantic meaning. In addition, sounds made by the artist which are not intended to convey meaning, but instead harmonize or contribute to the melody also ended up in the data like the string 'ladedadada'. For the purposes of this paper, these strings were not removed because they did not significantly hinder the aims of the paper and would have been very difficult to identify simply.

3 Preparing the Data

To prepare the data, the data had to be split and preprocessed in several ways. In this experiment, it is split into two Jupyter notebook files which can be found in this project's GitHub repository[5]. First, a sample of the data was taken from the full data. This was done due to computational limitations meaning that more data could not be used. Next, the data was trimmed to only contain songs labeled as being in English. After that the data was saved into two separate files for the top two genres represented in the data: pop and rap.

After the data was separated, it needed to be both cleaned and put into a single text file. The dataframe was read for each genre, and line by line, the data was cleaned to remove problematic punctuation and line breaks. A potential source of error comes from how the data was split at line breaks. In a song, most sentences are spread across several lines, and the punctuation for the end of a sentence may not be clear. For this reason, new lines were simply treated as spaces with the hope that most songs would provide punctuation at the end of a line if it was the end of a sentence. This was not always true, but worked better than treating each line as a sentence which did not provide a large enough context window for the model. Every line from every song in the category was added to a larger text file with only lyrics and no other data. These text files were then ready for the embedding step.

4 Training the Embedding Spaces

Once the data was prepared, embeddings were trained using the Gensim Word2Vec package. After the data was vectorized according to sentences and words, the data had to be further reduced in size by a factor of ten due to computational limitations. Around 315 and 375 thousand sentences were used to train the embedding space for pop and rap respectively after the data being trimmed.

The embeddings were created using a skipgram model because, when compared to a CBOW approach, the skipgram model seems to provide more insight into semantic meaning of the data which aligns more closely with the goals of this paper. After the model was trained, it was saved for later use and data collection.

In the final step, twenty random words that were present in both the pop and rap embedding spaces were chosen for comparison. To compare the data, the twenty closest neighbors of each word in the pop/rap embeddings were collected and compared using a Jaccard similarity measure. This would provide data on how the semantic meaning of words differs in the pop music domain compared to the rap genre.

The implementation of this and details on how to perform it independently are on the GitHub repository for this paper[6]. Most of the word done to create the embeddings is encapsulated in the imported Gensim code and the rest of the work was primarily focused on processing, trimming, and comparing the data.

[5] https://github.com/JTSIV1/Pop-and-Rap-Music-Embedding-Space-Comparison/tree/main

[6] https://github.com/JTSIV1/Pop-and-Rap-Music-Embedding-Space-Comparison/tree/main

5 Results and Evaluation

Looking at the Jaccard similarity measures, they are low. The mean across the twenty randomly sampled words is 0.0976. This equates to an average of 1.8 words of overlap in the top twenty closest words to the sample word in the two embedding spaces. Initially, this would suggest that the two genres are very different linguistic domains.

Below in figure 1 is a table of the Jaccard similarity for 20 randomly sampled words.

Word	Jaccard Similarity
‘day’	0.111
‘go;’	0.000
‘rock’	0.000
‘ten’	0.538
‘dreamin”	0.250
‘verse’	0.379
‘10:30’	0.000
‘dolet’s’	0.000
‘rogan’	0.000
‘war’	0.081
‘rise’	0.053
‘yeahh’	0.026
‘1b’	0.000
‘breath’	0.143
‘blowfly:’	0.000
‘fuck’	0.053
‘pete’s’	0.026
‘44th’	0.000
‘trapped’	0.212
‘8x’	0.081
Average	0.098

Figure 1: Jaccard Similarity for 20 Randomly Samples Words in Rap and Pop Embedding Spaces

First, the closest neighbors were analyzed to determine if the results of the embedding training were reasonable. Looking at the word ‘breath’ for example, the top four neighbors for pop and rap were as follows in figure 2.

Pop	Rap
‘gasp’	’breaths’
‘breaths’	‘gasp’
‘gasping’	‘gaspin’
‘lungs’	‘breathe’

Figure 2: Top four Neighbors for ’breath’ in Pop and Rap Embedding Spaces

All of these are semantically similar words and are similar in the two categories. Sill, even though it would seem the embedding spaced are similar, the Jaccard similarity score for breath was 0.143, which is low. This is because the words in the data may not have gone through enough preprocessing. For example, the plural “breaths” should not be in the data, and variations on the word “gasp” like “gasps”, “gasping”, “gaspin’”, and “gaspin” were all present in the data when they should not have been separate entries. This may have led to a lower similarity, because the same word functionally has several embeddings which may not each individually have enough data to end up in the exact correct spot. This both clutters the top 20 of each with different forms of the words leading to fewer potentially overlapping words, but also makes it less likely to actually find overlap with the repeated word[7].

In the random selection of words, an interesting discovery came from swear words. The embedding space for each genre provided useful information about the genres from what the nearest neighbors were. See the top four words for each neighborhood in figure 3.

Pop	Rap
‘f*ck’	’motherfuck’
‘shit’	‘fucks’
‘f**ked’	‘muthafuck’
‘navarro’	‘fucc’

Figure 3: Top four Neighbors for ’fuck’ in Pop and Rap Embedding Spaces

Looking at swear words, the pop category had neighbors that were also swear words, but with asterisks for censoring the word, or more mild swear words in the same category. On the other hand the rap data had no censorship, and many examples of swear words spelled in more colloquial ways or with letters repeated. This shows that in the rap domain swearing is not restricted in the same way that it is in the pop music domain, and that it is more fundamental to the genre and even has generated slang versions of the given harsh language.

Another interesting observation comes from metadata that for whatever reason did not get automatically removed. In the embedding spaces, there are embeddings for strings like ’8x’ indicating that the next segment of the lyrics is to be repeated eight times. In the random sample of words collected, ’8x’ was among the embeddings.

In figure 4, there is a clear difference between the neighbors in the pop space versus the rap space. In pop songs, the metadata is much more likely to occur with strings that are used to harmonize or create melody in the song. By contrast, in the rap embedding space the neighbors are mostly other

[7]It is possible for two neighborhoods to have the same root but each have a different form of the word, causing the similarity to be lower as they would not count as an intersection

Pop	Rap
'*repeat'	'4x'
'whoaaa'	'2x'
'fabolous'	'10x'
'woooh'	'repeat'

Figure 4: Top four Neighbors for '8x' in Pop and Rap Embedding Spaces

metadata instructions. This is likely because in pop harmonies are often repeated, and the same sound may be written in several songs. With rap however, there are fewer melodic additions like that and instead the metadata would just be placed with more lines of lyrical content which are more likely to be unique across songs.

6 Future Explorations

Looking into the future, there are three broad possibilities for further research: perfecting semantic comparison for the genres; looking into syntactic comparisons of the genres; or exploring genres outside of pop and rap.

The approaches taken in this paper provide valuable semantic insights and a good foundation for comparison, however they do not provide a perfect semantic analysis. Due to computing limitations, only a small portion of the available corpus was used in training the embedding spaces used in this paper. Half of the original corpus was loaded, and later on 90% of what was loaded was thrown away. This means that 90% of the corpus for pop and rap lyrics was left unused. For an improved model, the entire corpus should be used and trained for the proper amount of time. In addition, issues with pre-processing should be resolved to ensure that all forms of the same word end up represented in the same embedding. This would provide a more accurate and informative analysis, putting more weight behind the Jaccard similarity scores, and reducing the number of nonsense words in the embedding space.

Another approach could be taken in comparing rap and pop, and instead of looking at meaning, an experiment could look at structure. Music in general uses a different syntactic register than everyday spoken or written varieties, and rap and pop would likely have very distinct syntactic norms. Using a model like CBOW for training the embeddings instead of Skip-gram could lead to more syntactic information being included in the embeddings. The CBOW approach is known to be a better approach for syntax, so its use may provide interesting introspection into the grammar systems in each musical genre.

Lastly, more work could be done to compare genres outside of pop and rock. These genres were chosen for this paper simply because they are the two most common in the corpus from Genius, however, there are many more genres with their own distinct styles and impacts that could be compared like rock for example. There is more in the Genius corpus that could support that research, and many other corpora exist with lyrical information encoded for different genres that could be used.

7 Conclusions

This paper explored semantic aspects of language found in pop and rap music lyrics, using a corpus collected from Genius. Through the use of a skip-gram model approach, embedding spaces were created for both genres, leading to a comparison of their semantic patterns. In spite of challenges such as data formatting issues and stylistic variations in the lyrics, valuable insights were found in the analysis section of the paper.

The results indicated that there are discernible differences between pop and rap music in their semantic meaning, as shown by the low Jaccard similarity scores in the comparison of the embedding spaces. While the embeddings provided reasonable representations of semantic similarities within each genre, there still exists significant room for improvement in pre-processing steps and data use to increase the accuracy and strength of the analysis.

In addition, the paper suggests more opportunities for future research, including improving semantic comparison methods, exploring syntactic differences between pop and rap, and extending the analysis to other genres beyond pop and rap. By researching and exploring these areas, a more comprehensive understanding of the linguistic characteristics of lyrics in different music genres is possible.

Overall, this paper contributes to the ongoing discourse on the intersection of language and music, offering insights into how semantic patterns appear in the lyrics of pop and rap songs. As technology and methodologies continue to evolve, further investigations into this intriguing field will produce deeper and more concrete insights into the diversity of human expression through music.