Analysis Of Emotional and Psychological Themes in Music over time and across genres

Harsh Ramani, Dhruv Charan, and Aryan Patil

Introduction:

Our application delves into the profound connection between human emotions and music, aiming to unveil the sentimental and thematic essence embedded within song lyrics. By deciphering these intricacies, we offer insights into the cultural significance of music as a medium of self-expression and emotional communication. This exploration is vital for individuals interested in understanding the psychological impact of music and its role in shaping societal narratives. Through our project, we shed light on the intersection of linguistics, psychology, and musicology, making it relevant and compelling for researchers, music enthusiasts, and those intrigued by the human experience.

In our endeavor, we tackle various domains of Natural Language Processing (NLP), including sentiment analysis, trend analysis, and lyric generation. Leveraging cutting-edge NLP models such as BERT, XLNet, and a variant of GPT-2, we analyze, classify, and generate text to unravel the complexities of lyrical content across different genres and eras. Our approach encompasses classification, topic modeling, and language generation, providing a comprehensive examination of linguistic analysis within the context of music. Through these methodologies, we aim to offer nuanced insights into the emotional and thematic landscape of music lyrics, enriching our understanding of cultural expression and human sentiment.

*Research Goal:*Our research goal is to address the issue of the evolving music landscape where certain genres may face dismissal due to shifts in popularity over the decades. We propose a solution that focuses on the evolution of emotional themes in music, rather than the genres themselves.

Background:

Inspired by Glenn McDonald's 'Every Noise at Once' project at Spotify, which created a musical-genre space for user exploration, our aim is to delve into lyrical content for enhanced music discovery. Bonta and Venkateswarlu explored sentiment analysis techniques applied to music reviews, aiming to understand the sentiment expressed towards various songs or albums. Their work provides valuable insights into how listeners perceive and interpret music[1]. In a similar vein, Kuo and Chiang focused on recommending music based on emotions. They analyzed associations between music and emotions in film soundtracks[2].In a comprehensive review of research on emotion in music, Juslin and Sloboda covered topics such as emotion induction, perception of emotion, and the role of music in emotional expression[3]. Finally, Napier and Shamir's study aimed to assist songwriters in crafting lyrics that evoke specific emotional responses, demonstrating the practical applications of sentiment analysis in music[4].

*Drawbacks:*These researches provide significant understanding of sentiment analysis in music, emotion detection, and the convergence of Natural Language Processing (NLP) with music examination. Yet, they overlook the creation of music suggestion systems or lyrics production based on psycholinguistics.

Dataset:

Our project utilizes two datasets from the Hugging Face library. The first, chloeliu/lyrics[5] contains 28,372 entries featuring stemmed song lyrics alongside artist information, genre, release date, and musical characteristics like acoustic ness and instrumentalness. This dataset allows us to explore thematic content, emotional tone, and musical aspects of songs. We'll split this data, using 25,000 entries for training and the

remaining 3,372 for testing tasks like sentiment classification and genre prediction. The second dataset, "amzar1303/lyrics," focuses on lyric generation. It contains 4529 records with information like track title, lyrics, artist name, and URLs. Here, we'll use 80% of the data for training our lyrics generation model and the remaining 20% for validation purposes[6].

Dataset	Labels	Avg song length	Min length	Max length	Key Words
chloeliu	Acoustics Spiritual Feelings	446	4	1714	Think, Know, Life
amzar1303	NA	1219	27	6392	Love, Want, Know

Table: Important statistics of the datasets

Methods:

Sentiment Classification:
This study tackles sentiment classification**(I. Syntax|Classification)** using BERT and XLNet, two powerful **Transformer models(III. LM | Transformers)** known for their exceptional grasp of textual context and meaning, crucial for accurate sentiment analysis To harness these strengths, we fine-tuned both models with a specific configuration. The training process involved fine-tuning for 3 epochs to balance underfitting and overfitting. We used a batch size of 8 for optimal GPU memory usage and incorporated a 500-step warm-up phase to gradually adjust learning rates and stabilize the learning curve. The HuggingFace Trainer API streamlined training and performance tracking for both models.

To improve sentiment analysis, we created a custom model, "BertWithAdditionalFeatures." This model extends the pre-trained BERT by incorporating non-text features from the dataset, such as musical characteristics (acousticness, energy) and thematic elements (genre). These features provide additional context beyond lyrics.

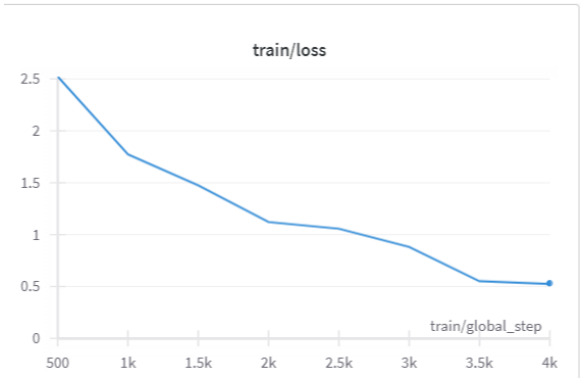

Fig: Train Loss for Xlnet model

Before integration, the non-text features are processed through a linear layer to reduce their dimensionality and facilitate smoother combination with textual features. The processed features are then concatenated with BERT's pooled output, which captures contextual information from the lyrics. This combined representation is fed into a dropout layer and a final classifier.This hybrid approach leverages both BERT's strength in understanding language context and the additional context provided by non-text features. The goal is to achieve a more nuanced sentiment analysis that goes beyond the surface meaning of the lyrics.

Lyrics Generation:
Our study incorporated a fine-tuned GPT-2**(III. LM | Transformers)**variant as a baseline for lyric generation**(III.Application)**. This involved training the model on 80% of the available data (4,000 samples). To optimize its performance for this task, we configured the model with specific parameters that influence its understanding of lyric structure and its ability to generate coherent sequences. These parameters included output hidden states, layer normalization, dropout rates, and attention mechanisms.The training process itself utilized the Adam optimizer with a learning rate of 1e-6 and spanned three epochs. Perplexity served as the evaluation metric, gauging the model's effectiveness in predicting the next word in a sequence. To further enhance the model's lyric generation capabilities, we implemented several key adjustments. Disabling caching ensured all computations during training remained up-to-date. Additionally, layer normalization and dropout were employed to

improve the model's overall performance. Finally, we set the number of classes to 1, signifying its use for regression or single-class classification tasks. This configuration aligns the model's output more effectively with lyric generation.Following fine-tuning, we developed a function that enables lyric generation based on provided prompts. This function serves as a demonstration of the model's ability to produce meaningful and contextually relevant text.

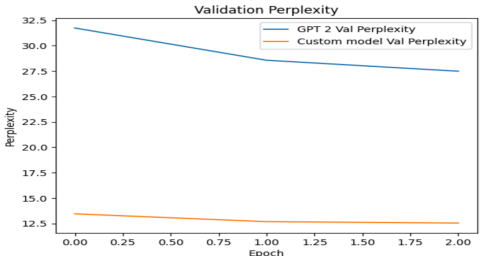

Fig: Train Loss and the Validation perplexity of GPT2 variant

Model Card: Lyrics Generation

Model Name:

Fine-tuned GPT-2[7](Custom Variant)

Fine-tuned Parameters: Output Hidden States (True), Use Cache (False), Layer Norm Epsilon (1e-5), Dropout (0.1)

Intended Use:

The objective of this model is to generate song lyrics given a prompt. This model is intended to be used for generating song lyrics for creative and entertainment purposes. It can be utilized by songwriters, musicians, etc.

Limitations and Ethical Considerations:

1. Language Bias: The model's outputs may reflect biases present in the training data, including cultural, gender, or racial biases.

*2. Originality:*The generated lyrics may resemble existing songs or lyrics due to the nature of the training data and the model's ability to learn patterns.

Evaluation Metrics:

Perplexity: A measure of the model's uncertainty in predicting the next token in a sequence.

Human Evaluation: Subjective assessment by human evaluators to judge the coherence, relevance, and creativity of the generated lyrics.

Training and Evaluation Data:

Training Data: The model was fine-tuned on a custom dataset comprising song lyrics collected from various sources.

The dataset includes lyrics from diverse genres and artists to ensure a broad representation of language styles and themes. The training set consists of 80% of the total data, randomly selected.

Validation Data: A randomly chosen 20% subset of the training data was used for model evaluation and validation.

Caveats and Recommendations:

The generated lyrics from this model should be interpreted with caution due to potential contextual limitations, biases inherited from the training data, and vocabulary constraints. Users are advised to provide clear and relevant prompts.

Fig: Word Cloud for lyrics on new data

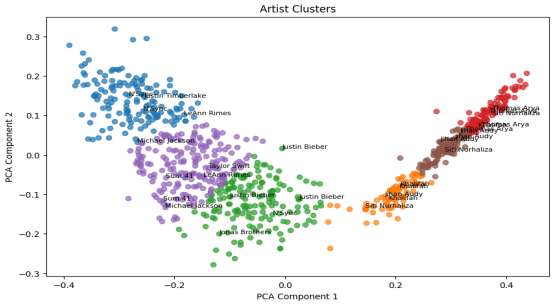

Fig: PCA based biplot for Artist

L*DA based Recommendation system(Variant):*
Our song recommendation system leverages both textual and numerical features from the "chloeliu/lyrics" dataset. This dataset includes song lyrics alongside numeric attributes like thematic categories ('dating', 'violence'). The text processing stage involves transforming lyrics into a matrix using CountVectorizer. This removes stop words, considers word frequency within documents and across the dataset, and limits features to a manageable number. LatentDirichletAllocation (LDA) is then applied to this matrix, a technique that assigns each song to a set of 10 pre-defined topics (a tunable hyperparameter in our case) to perform **Topic Modeling(II. Semantics | Probabilistic)**

Next, we combine the topic distribution from LDA with selected numeric features from the dataset. These numeric features are normalized for consistency with the LDA output using MinMaxScaler. Cosine similarity is then calculated between all song pairs based on the combined features, resulting in a matrix of similarity scores. Finally, an **application of NLP(IV. Application)** these similarity scores to recommend songs similar to a provided input song. It takes a song index and a desired number of recommendations as input, returning the names of the most relevant songs.

Evaluations:
Sentiment Classification:
We evaluated the performance of three models for sentiment generation: BERT, XLNet, and a variant called BERT with Factor Adaptation (BERT-FA). Accuracy and F1-score were used as our evaluation metrics, with the results presented in a table in the next section.
Among these models, BERT-FA achieved the best performance, although by a slight margin compared to standard BERT and XLNet. It's worth noting that BERT-FA did not utilize any categorical features from the dataset, limiting the information it could process. Despite this limitation, it still delivered good results. This suggests there's potential for improvement in BERT-FA through a more robust architecture. This new architecture would need to effectively handle non-text features and integrate them seamlessly with the outputs from the BERT.

Another possible avenue for improvement lies in exploring alternative mechanisms that require fewer resources. This could involve freezing the BERT layer, extracting penultimate embeddings (representations from the second-to-last layer), and then adding a simpler architecture for transfer learning over the dataset. While we weren't able to implement this approach in our current work, it remains a promising direction for future exploration.

Lyric Generation:
Our fine-tuned GPT-2 variant significantly outperformed the baseline model in perplexity, achieving a score of 12.48 compared to 35.23. This improvement likely stems from adjustments to the model architecture, particularly layer normalization, dropout rate, and potentially other hyperparameters. These refinements enabled the variant to better grasp the patterns within the lyric dataset, resulting in lower perplexity.

The quality of generated lyrics further reinforces this conclusion. The baseline model produces repetitive and shallow lyrics lacking coherence. In contrast, the variant generates lyrics with improved coherence, semantic relevance, and creativity. This suggests that fine-tuning has enhanced the model's ability to understand and produce contextually meaningful text. Overall, the perplexity reduction and qualitative improvements demonstrate the effectiveness of the modifications made to the GPT-2 variant.

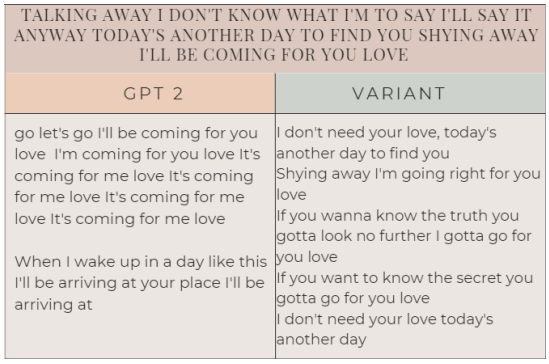

Fig: Comparing the two models for generation

LDA based Recommendation system(Variant):
We compare the traditional LDA with its variant where we have integrated numeric features like

'sadness', 'feelings', 'danceability', 'loudness', 'acousticness', 'instrumentalness', 'valence', 'energy', etc to recommend songs. The main assumptions made in this process are that the lyrics and the selected numeric features are representative of the songs and can be used to calculate meaningful similarity scores, and this was proved to be true using hypothesis testing(T test and KS test) with significantly lower p values(<<0.05). However, the actual results can be analyzed only after testing the recommendations based on actual users.

Liked Song: Blood Beach	
LDA without Numeric Features	LDA integrated with Numeric Features
the moron brothers hippa to da hoppa i'm blue blue money red light	the staircase (mystery) night shift Mercy poptones like an outlaw (for you)

Table: Comparison of recommendation systems

Task	Model	Evaluation	
		Accuracy	F1
Sentiment Analysis	BERT	62%	0.62
	XLNET	59%	0.6
	BERT Variant	65%	0.63
		Perplexity	
Lyrics Generation	GPT2	35.93	
	GPT2 Variant	12.48	
		Avg Sim Score	
Recommendation System	LDA	0.28	
	LDA Variant	0.65	

Table: Evaluation of all the models

Conclusions:

We implemented a novel approach for sentiment analysis and Recommender System using Human Factors. In conclusion, our project has explored the evolution of music over the decades, focusing on the rise and fall of various genres. We identified a potential issue where certain genres might be dismissed or recede in popularity over time. However, our research suggests that the emotional themes in music have evolved and remained consistent, regardless of genre. Therefore, we propose that the solution lies not in focusing on genres, but rather on these emotional themes. By incorporating these themes into music, we can ensure the continued relevance and survival of all genres. This approach recognizes the inherent value and unique contributions of each genre, and promotes diversity and innovation in music.

References

[1]Bonta, Venkateswarlu & Kumaresh, Nandhini & Naulegari, Janardhan. (2019). A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis. Asian Journal of Computer Science and Technology. 8. 1-6. 10.51983/ajcst-2019.8.S2.2037.

[2]Fang-Fei Kuo, Meng-Fen Chiang, Man-Kwan Shan, and Suh-Yin Lee. 2005. Emotion-based music recommendation by association discovery from film music. In Proceedings of the 13th annual ACM international conference on Multimedia (MULTIMEDIA '05). Association for Computing Machinery, New York, NY, USA, 507–510. https://doi.org/10.1145/1101149.1101263

[3]Medinic-Kazazic, M. (2012). Handbook of music and emotion: Theory, research, applications: Canadian journal of music. Intersections, 33(1), 107-110,116.

[4]https://doi.org/10.1525/jpms.2018.300411

[5]https://huggingface.co/datasets/chloeliu/lyrics

[6]https://huggingface.co/datasets/amzar1303/lyrics/viewer/default/train

[7]Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multi task learners. OpenAI Blog, 1(8), 9.