Analysing the Sentiment of British Top of the Pops Programmatically - 1961 to 2024

Theodore Proctor
Department of Computer Science
University of Exeter, Exeter, UK

(Internal or External) Supervision:
Name of supervisor
Supervisor affiliation

(Internal or External) Supervision:
Name of supervisor
Supervisor affiliation

***Abstract*—This research applies Sentiment Analysis to British Top of the Pops lyrics (1961-2000), exploring the connection between music sentiment, societal changes, and historical events to understand the era's cultural landscape.**

I. Introduction:

The allure of music lies not only in its rhythm and melody but also in its ability to evoke emotion. We explore British Top of the Pops from the 1970s through to the 2000s, a period that witnessed some of the most influential changes in music history [1], in part due to globalisation, resulting in the creation of new genres, such as rock and punk.

In order to study music, we will be applying NLP (Natural Language Processing) [2] techniques, such as SA (Sentiment Analysis). SA is the interpretation and classification of emotions within text data [3]. When applied to music, it becomes a tool to analyse the meaning of lyrics, offering insights into the mood and emotions conveyed by the artists [4]. This is particularly useful when examining large quantities of lyrics, allowing us to uncover and recognise trends.

This project aims to discover such trends. Due to the influence of music on culture [5], we endeavour to combine this with societal fads and events to investigate correlation or causality. We also hope to map the emotional landscape of music over the ages.

We begin by recognising music as a reflection of the times, a mirror to the zeitgeist of an era. SA allows us to map the emotional undercurrents of British pop and rock against the backdrop of societal changes. This offers insights into the genre's cultural impact, showing how it both reflects and influences its historical context.

II. Motivation:

The investigation into the changing sentiment of British rock across several decades probes the intersection between cultural trends and societal moods, hypothesising that lyrical content serves as a barometer for historical and social fluctuations. This provides valuable insight into decades past.

Although quantitative studies into pop music have been completed [6], none have narrowed the scope of research to British pop and rock across 1961-2000, normally reflecting on the industry or era as a whole or studying the evolution of genres [7].

The global impact of British pop and rock [8] renders it a subject worthy of attention. By studying its lyrics using NLP techniques such as SA, we aim to delineate the historic emotional trajectory of the nation, offering insights into the United Kingdom's culture over time. This can contribute to a deeper understanding of how cultural outputs are interwoven with historical and social contexts, and can be expanded to a global scale. Furthermore, this should also allow for an expansion on the psychology behind music [9] and how it affects emotions.

One example that exemplifies affect on emotions is the song "Bangla-desh" by George Harrison [10]. This song, created as a means to raise awareness about the crisis in Bangladesh [11], impacted many listeners hearts, and resulted in a great appeal. Despite the surrounding media attention, which no doubt improved the reception of the song, the remorseful and pleading tone of the music and its lyrics embodied the cry for help, and managed to positively impact the Bangladeshi people [12].

Another study that may prove fruitful is the gender of the lyrical subject. This allows us to consider the gender norms of the zeitgeist and perhaps determine whom the music is catered towards, perhaps elaborating on the media's outlook on gender and stereotyping.

This report thus seeks to address a significant gap in existing research by providing a quantified sentiment analysis of British rock lyrics, offering a novel contribution to the fields of musicology, cultural studies, and psychology.

III. Aims:

- To map the sentiment trajectory in British music from the 1970s to the 2000s.
- To provide a quantitative analysis of British music over time.
- To interpret the emotional tone of lyrics in the context of societal and cultural dynamics.
- To correlate lyrical sentiment with key historical events and cultural shifts.
- To observe trends in the gender of the subject of a song.

IV. Objectives:

- Acquire a corpus of lyrics from 1961-2000.
- Apply NLP techniques to the lyrics of top British songs for SA.
- To categorise sentiment scores for the lyrics corpus.
- To interpret the results, investigating emotional undertones and changes the music.
- to compare findings with historical literature and events.

V. Success Criteria:

- Analyse sentiment in the top 100 British songs yearly from 1961-2000.
- Validate the transformers model accuracy with proved methods like lexicon-based.
- Acquire and clean a comprehensive lyrical corpus.
- Link lyrical sentiment trends with historical and cultural events.

VI. Technical Solution - Dataset and Acquisition:

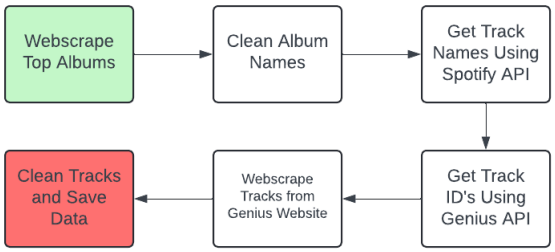

Fig. 1: Data Acquisition Flowchart

We decided on this time frame (1961-2000), since it captures cultural and musical shifts that occurred over the recent decades. Despite collecting a broader range of data, we sample the top albums each year 100 times (randomly) for analysis. This is due to the limited scope of the project and the data loss that occurred, with some months of data being unavailable.

To acquire our dataset of British Top of the Pops from 1961-2000, a systematic approach using a combination of web scraping, NLP, and data storage techniques was utilised. The flow of data acquisition, as detailed in 1, started by web-scraping the Official Charts [13] website, which is a "trusted weekly measure of what is popular in music right now", specifically for British top charts. Once the top albums of each week have been acquired, Spotify API [14] is used to extract the tracks from each album, before Genius API [15] is used to get track IDs. Finally, we web scrape the Genius website, extracting the lyrics from the page associated with the track ID. API requests were handled using Pythons requests library [16], and all web scraping utilised requests and Beautiful Soup 4 [17].

Throughout this process, we systematically saved the phases of data collection. This was primarily done using a file structure, saving datums as .json files, and lists as .csv files. Entries nomenclature used "year" and "month" elements to simplify data retrieval. Additionally, we applied error handling, which was used when an API call or web scrape operation returned null. These decisions were made due to the inherent instability of web-based operations.

During data collection, we had to ensure that any strings being passed were in the correct format, with all escape characters removed. Once the data has been acquired, it undergoes pre-processing to ensure its quality before analysis. This includes removing stop words, punctuation, and noise, as well as normalising the case and applying stemming or lemmatization techniques to reduce words to their root forms.

A. Ethics and Risk Management

Web scraping is performed respecting website terms of service, considering the website's ROBOTS.txt. Privacy concerns are addressed by using only publicly available lyrics, with no personal data involved. Risk mitigation includes adhering to fair use standards, using public APIs with proper agreements and other publicly available data.

VII. Technical Solution - Analytical Techniques:

We will analyse these lyrics by a plethora of means. This will allow for a comprehensive investigation, and should allow one to examine any trends and correlations present with greater nuance.

A. Sentiment Analysis - Lexicon-Based

VADER [18], employed for lexicon-based sentiment analysis, excels in analysing more informal text sentiments. This is a valuable consideration when applying SA to expressive content like song lyrics. It assigns sentiment scores (positive, negative, or neutral), enabling efficient sentiment trend identification in our lyrical data. Lexicon-based sentiment analysis models rely on a predefined dictionary of words with associated sentiment values to assess the emotional tone of text.

B. Sentiment Analysis - Transformer-Based

A sentiment analysis transformer model [19] is an advanced machine learning tool that uses transformer architecture [20]. Known for its efficiency in handling sequential data, it can produce more nuanced classifications. For a more nuanced sentiment analysis, we utilise a transformer-based approach using Hugging Face's [21] pipeline with the "SamLowe/roberta-base-go_emotions model" [22]. This model is adept at discerning a broader spectrum of emotions, categorising text into one of 28 distinct emotional states.

C. Gender Subject Over Time

The temporal evolution of gender portrayal in song lyrics is examined using predefined sets of male and female-associated terms. By quantifying the occurrence rate of these gender-specific terms, we gain insights into the changing representation and focus on male and female subjects over different time periods. This aspect of the study provides a perspective on the dynamics of gender representation in music.

D. Lexical Richness

Lexical richness [23] in our study is quantified by measuring the diversity of vocabulary used in the lyrics, providing insights into the creativity and complexity of the lyrical content. It is evaluated using the formula $L = \frac{\text{unique words}}{\text{total words}}$, where L represents lexical richness.

VIII. Technical Solution - Conclusion:

For our final analysis, we will employ a transformer-based VII-B approach, using a lexicon-based VII-A method as validation. Statistical tests (such as t-tests) can be used for this purpose, comparing the lexicon-based sentiment value with an aggregated value from the transformer-based model. Both lexical richness VII-D and gender subject VII-C will be used as secondary measures.

IX. Results:

The figures 2 3 present the findings of text analysis applied to British Top of the Pops from 1961 to 2000.

Given the nuance provided by the transformer method VII-B, we introduced aggregation of the positive and negative emotions, so that our results could be plotted. Subsequently, we introduced a normalisation factor to ensure comparability across the dataset. Additionally, we adopted a dual-validation method where an emotion is considered for analysis only if both employed SA techniques concur in their assessment. This consensus requirement reinforces our confidence in the emotional trends reported. The backdrop of Fig. 2 is colour-coded, where each year's dominant emotions (excluding neutrality)

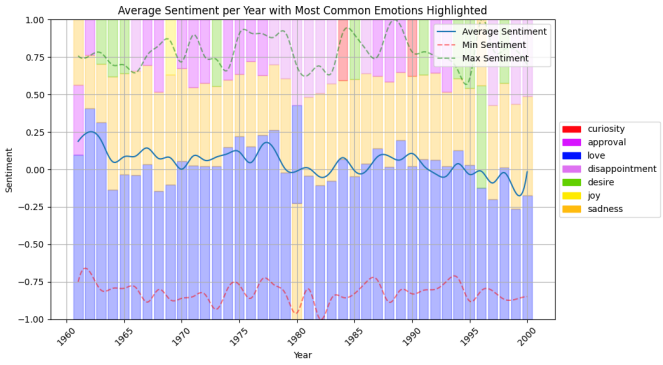

Fig. 2: Sentiment Analysis

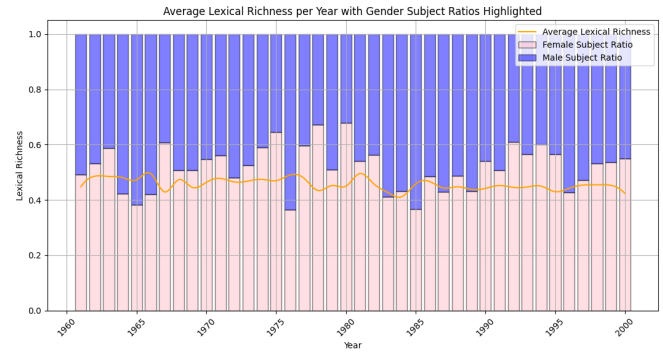

Fig. 3: Lexical Richness and Gender Subject

are visually highlighted. The top three emotions for each year are represented by distinct background colours.

Fig. 3 contains a line chart, displaying lexical richness over time, as well as the ratio of gender subject as the backdrop.

A. Sentiment Analysis Discussion

From observing Fig. 2, we can observe a steady decline in sentiment, along with predominant emotions being love, sadness, and disappointment, aligning with the prevalent theme of love in songs.

The sentiment reduction may indicate a trend towards more sombre or intricate themes. Given that the data includes multiple genres, it might reflect changing genre popularity, for instance, a rise in genres with typically lower sentiment scores like punk, grunge, or certain types of electronic music [24].

Over time, one can notice an increase of the emotions desire and curiosity, which may suggest a societal trend towards freer emotional expression and challenging norms. These changes align with a decline in approval, thus corresponding with music evolving towards more critical or unconventional perspectives, consistent with wider societal changes.

The only year where joy was in the top three emotions was 1969, perhaps due to major events like the moon landing and Woodstock. Around 1980, a significant sentiment fall coincided with Thatcher's tenure, possibly reflecting the UK's political and social environment in music. A noticeable sentiment decline occurred in 1995. This period saw the rise of Britpop, a movement that brought bands like Oasis and Blur into the spotlight, their music often characterised by a mix of melancholic and introspective themes. The late 1990s exhibited a notable trend: a sharp sentiment decline followed by a spike in 2000, potentially reflecting the turn of the millennium's contemplative and forward-looking nature in music.

B. Discussion of Lexical Richness

During the 1970-1975 period, an upsurge in sentiment, coupled with consistent lexical richness, may correlate with the emergence of more optimistic music genres such as disco and soul, as well as the lingering influence of the peace movement.

After 1980, a notable year for low sentiment, there was a peak in lexical richness. The lack of a discernible pattern between sentiment and lexical richness suggests that the complexity of language in songs does not directly relate to the emotional tone of the lyrics. This indicates that while the vocabulary used in music might become more sophisticated or varied, it does not necessarily reflect a corresponding change in emotional content.

C. Discussion of Gender Subject Over Time

The trend towards gender balance in pop music, highlighted by artists like Madonna and Kate Bush, mirrors broader gender equality movements. After 1980, a decrease in female-focused songs possibly indicates a thematic shift or counter-movement. The rise in female-centric songs until 1983 aligns with the 1970s women's liberation movement's impact on music.

However, the lack of a strong trend accompanying the rise of female pop stars could imply that this genre was more influential or evolving differently in another market, such as the United States.

The period between 1975 and 1980 saw a peak in gender subject variance, possibly due to more prominent female artists or gender-focused themes in music. However, this could be attributed to the limitations of the dataset.

Overall, the analysis indicates that variations in gender focus within songs and sentiment levels are not directly linked, implying that the presence of male or female subjects doesn't necessarily govern the overall sentiment of the music.

X. Conclusions:

From analysing the British Top of the Pops, it seems as though neither gender focus nor lexical richness are directly linked to sentiment. A pertinent note is that correlation does not imply causation. Cultural, social, and political factors play significant roles in influencing music, and these findings could vary greatly by region and genre. To confirm any correlations, a statistical analysis would be necessary. It's also important to consider the limitations of SA and other NLP techniques, which might not fully capture nuanced expressions in music, especially when considering the instrumental layers.

A. Future Works

Further works include enhancing the dataset by increasing its granularity (sampling by month), and applying statistical techniques to authenticate the observed correlation. Furthermore, refining the approach or model to more accurately detect irony and other linguistic nuances could improve its efficacy. Extending the scope of the study to the present day could be interesting, due to the introduction of genres like grime and drill music. Finally, conducting a comparative analysis with other art forms and the relationship with the revenue of the UK music industry would be interesting developments.

REFERENCES

[1] M. Verboord and A. Brandellero, "The globalization of popular music, 1960-2010: a multilevel analysis of music flows," *Communication Research*, vol. 45, no. 4, pp. 603–627, 2018.

[2] J. P. Mahedero, Á. Martínez, P. Cano, M. Koppenberger, and F. Gouyon, "Natural language processing of lyrics," in *Proceedings of the 13th annual ACM international conference on Multimedia*. ACM, November 2005, pp. 475–478.

[3] T. Nasukawa and J. Yi, "Sentiment analysis: Capturing favorability using natural language processing," in *Proceedings of the 2nd international conference on Knowledge capture*. ACM, October 2003, pp. 70–77.

[4] "How art works," Podcast, 2018, duration: 10 min 18 sec.

[5] A. Bennett *et al.*, *Popular music and youth culture: Music, identity and place*. Macmillan Press Ltd., 2000.

[6] K. Napier and L. Shamir, "Quantitative sentiment analysis of lyrics in popular music," *Journal of Popular Music Studies*, vol. 30, no. 4, pp. 161–176, 2018.

[7] S. Kamalnathan, Y. Mishra, V. Kumawat, and V. Bangwal, "Evolution of different music genres," 2019.

[8] D. Simonelli, *Working Class Heroes: Rock Music and British Society in the 1960s and 1970s*. Lexington Books, 2012.

[9] D. Deutsch, "Psychology and music," in *Psychology and its allied disciplines*. Psychology Press, 2019, pp. 155–194.

[10] G. Harrison, "Bangla-desh," Genius Lyrics, July 28, 1971, https://genius.com/George-harrison-bangla-desh-lyrics.

[11] "Bangladesh liberation war," last modified 2024, accessed March 31, 2024. [Online]. Available: https://en.wikipedia.org/wiki/Bangladesh_Liberation_War

[12] T. Report, "The concert for bangladesh: How music influenced the war," *The Business Standard*, December 2020. [Online]. Available: https://www.tbsnews.net/bangladesh/victory-day/concert-bangladesh-how-music-influenced-war-172534

[13] "Official singles chart top 40," Official Charts Company, March 2024. [Online]. Available: https://www.officialcharts.com/charts/uk-top-40-singles-chart/

[14] "Spotify web api," Spotify for Developers, 2024. [Online]. Available: https://developer.spotify.com/documentation/web-api/

[15] "Genius api documentation," Genius Developer Portal, 2024. [Online]. Available: https://docs.genius.com/

[16] "requests," PyPI, May 2023, version 2.31.0. [Online]. Available: https://pypi.org/project/requests/

[17] "beautifulsoup4," PyPI, January 2024, version 4.12.3. [Online]. Available: https://pypi.org/project/beautifulsoup4/

[18] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Proceedings of the international AAAI conference on web and social media*, vol. 8, no. 1, 2014, pp. 216–225.

[19] U. Naseem, I. Razzak, K. Musial, and M. Imran, "Transformer based deep intelligent contextual embedding for twitter sentiment analysis," *Future Generation Computer Systems*, vol. 113, pp. 58–69, 2020.

[20] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques," in *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*. Association for Computational Linguistics, Jul. 2002, pp. 79–86. [Online]. Available: https://aclanthology.org/W02-1011

[21] "Hugging face: The ai community building the future," Hugging Face, 2024. [Online]. Available: https://huggingface.co/

[22] S. Lowe, "roberta-base-go_emotions," Hugging Face Model Repository, 2024. [Online]. Available: https://huggingface.co/SamLowe/roberta-base-go_emotions

[23] B. Laufer and P. Nation, "Vocabulary size and use: Lexical richness in l2 written production," *Applied linguistics*, vol. 16, no. 3, pp. 307–322, 1995.

[24] P. Grossman, "Identity crisis: the dialectics of rock, punk, and grunge," *Berkeley journal of sociology*, vol. 41, pp. 19–40, 1996.