Lyric-based Playlist Continuation with Transformers

Suhas Hebbur Eshwar
shebbureshwar@ucsd.edu

Dhruthick Gowda Mohan
dgmohan@ucsd.edu

1 Introduction

Recommendation systems in any domain, be it related to videos, products, music, or news is fundamental in consuming large content effectively. With the remarkable surge in the number of subscribers to music streaming platforms, particularly during the Covid era (Szalai, 2020), the demand and significance for Music recommendation systems (MIR) have increased. However, majority of the systems have relied on acoustic features and social tags to identify and categorize emotions in music. Surprisingly, lyrics have been largely overlooked, despite their pivotal role in evoking emotions. Lyrics are semantically rich and expressive and have a profound impact on human perception of music (Ali, 2006). Despite a few studies showing superior performance with lyrics, the use of lyrics remains largely underappreciated. By harnessing the power of lyrics in recommendation systems, not only can users discover new music that aligns with their preferences and emotions, but it also allows for a deeper exploration of the artistic and emotional aspects of songs.

In this project, we aim to investigate the effectiveness of incorporating lyrics into a music recommendation system. We will extract meaningful information from lyrics using transformer-based approaches. Transformers have revolutionized natural language processing tasks by capturing the contextual relationships between words and generating high-quality representations of textual data. Given the sequential and contextual nature of lyrics, transformers are well-suited to capture the subtle connotations and emotional nuances expressed in song lyrics, enabling more accurate and fine-grained analysis for recommendation purposes. By integrating these techniques with existing music recommendation algorithms, we seek to create a more comprehensive and user-centric music recommendation system that takes into account the lyrical content of songs.

2 Related work

Music emotion classification using lyrics has been performed on traditional lexicons (Hu and Downie, 2010) (Hu et al., 2009). Not only do the lexicons possess a highly restricted vocabulary, but the values must also be combined without incorporating any contextual information.

Several previous studies have explored the recognition of emotions through the analysis of English song lyrics. The work by (An et al., 2017) utilized Naïve Bayes Classifiers to classify emotions into four distinct categories, (Happy, Angry, Relaxed, and Sad) achieving an accuracy rate of 68%.

In (Akella and Moh, 2019), multiple deep learning models were employed for mood classification using lyrics, including the Convolutional Neural Network (CNN), Bi-LSTM, and Convolutional Recurrent Neural Network (CRNN). The results indicated that the CNN model achieved an accuracy of 71%, while the Bi-LSTM model achieved 69.01% accuracy, and the CRNN model achieved 67.04% accuracy. However, in (Abdillah et al., 2020) the utilization of dropout parameters and activity regularization in the Bi-LSTM model yields a noteworthy accuracy of 91.08% in the task of emotion classification based on song lyrics. This achievement is particularly remarkable when compared to the performance of traditional machine learning methods in the same task.

When it comes to transformers, the work done by (Edmonds and Sedoc, 2021) reveals certain challenges associated with emotion classification of song lyrics using state-of-the-art techniques when utilizing out-of-domain data. BERT models trained on extensive collections of tweets and dialogue fail to generalize effectively to lyrical data, except for emotions related to joy and sadness. Conversely, models fine-tuned specifically on song lyrics achieve comparable levels of accuracy to models trained on out-of-domain data. Remarkably, these song-specific models maintain their performance even when working with significantly smaller lyrical datasets, aggregated from line to song level, annotated from diverse perspectives, and comprising various music genres. These findings emphasize the criticality of employing in-domain data for the accurate classification of emotions in song lyrics.

The study by (Agrawal et al., 2021), showcased the resilience of transformer-based methodology with fine-tuning for recognizing music emotions through lyrics across various datasets, surpassing previously employed approaches. In our work, we plan to employ a similar methodology to see how fine-tuned transformers on lyric data can help improve song recommendation for the task of playlist continuation.

The work by (Monti et al., 2018) took an ensemble approach by combining multiple recurrent neural networks, specifically Long-Short Term Memory (LSTM) cells, to predict the next track based on a sequence of tracks. By utilizing lyric metadata from the WASABI lyric corpus (Buffa et al., 2021), they developed lyric features that described different stylistic and linguistic dimensions of a song text, such as emotion and vocabulary. This highlights the importance of incorporating lyrics as a valuable source of information for understanding the content and emotional aspects of songs.

3 Your approach

For the task of playlist continuation, we plan to present a deep neural network structure capable of classifying songs into Emotion Quadrants (described in Section 4), as well as Valence and Arousal Hemispheres, based on given lyrics. We will try to fine-tune collectively on all these tasks using weight-sharing, which is a form of multi-task learning. By incorporating multi-task learning, we introduce an inductive bias that favors hypotheses capable of explaining all the tasks. The work done by (Agrawal et al., 2021) shows that this approach effectively addresses the concern of overfitting and diminishes the model's susceptibility to incorporating random noise during training, while also facilitating quicker convergence.

We will then use this fine-tuned transformer for the task of music recommendation, where we utilize the predictions made by the transformer jointly with user-playlist information to see if it results in better recommendation scores. For recommendation, we plan on using an LSTM-based recurrent neural architecture that takes in a sequence of tracks and associated information and outputs similar tracks for playlist continuation.

Baseline: We will use the Naive Bayes classifier as our baseline to classify a given lyric into one of 4 emotion classes - Happy, Sad, Relaxed, and Angry. The classifier will take in a vector representing the term frequency of each token in the lyric as input. We expect the classifier to perform decently, which can be further improved upon using the above-mentioned deep learning-based approach.

3.1 Schedule

We have divided our project completion timeline into the following subtasks, each of which we will be working on together.

1. Acquire and pre-process lyrics data (1 week)
2. Evaluate the baseline, fine-tune transformer architectures for Emotion Recognition and perform error analysis (2 weeks)
3. Pre-process playlist-lyric data for recommendation (1 week)
4. Build and evaluate the recommendation algorithm (2 weeks)
5. Work on the final report (1 week)

4 Data

MoodyLyrics (Çano, 2017): This dataset contains 2595 songs that are evenly distributed among the four quadrants of Russell's Valence-Arousal (V-A) circumplex model as shown in Figure 1 (Russell, 1980), which is a recognized framework for capturing musical emotions in a two-dimensional continuous space. In this model, emotions are represented as points in this space, where Valence indicates pleasantness and Arousal represents the energy content. To assign the V-A values at a word level, the authors utilized various existing lexicons such as ANEW, WordNet, and WordNet-Affect. These values were then averaged at the song level.

The validity of these assignments was further confirmed by incorporating subjective human judgment of mood tags from the AllMusic Dataset (Malheiro et al., 2018). To ensure high representativeness of each category, the authors only included songs in each quadrant that surpassed specific thresholds for both Valence and Arousal values.

Million Playlist Dataset (Chen et al., 2018a): This dataset, from the Spotify Million Playlist Dataset Challenge, is a research initiative aimed at facilitating advancements in music recommendations. It serves as an extension of the RecSys Challenge 2018 (Chen et al., 2018b), which took place from January to July 2018. The dataset provided includes 1,000,000 playlists, consisting of playlist titles and track titles, created by Spotify users between January 2010 and October 2017. The evaluation task focuses on automatic playlist continuation, wherein participants were required to predict the subsequent tracks in a playlist given a seed playlist title and/or an initial set of tracks.

Due to copyright restrictions, the provided datasets do not include lyrics. However, they do contain URLs from various lyric websites. To extract the lyrics, one possible approach is to create a web crawler for each website listed in the datasets. However, some of the URLs are broken, presenting a challenge. To overcome this issue, we propose a robust solution that involves leveraging the Genius website for extracting lyrics.

Existing APIs, including the Genius API, typically require accurate artist and track names to retrieve the lyrics. If there are misspellings in the dataset for either the artist or track names, the API fails to extract the lyrics. To address this problem, we implemented a web crawler that obtains the Genius website URL for the specific song, instead of relying on hard-coded artist and track names within the Genius API.

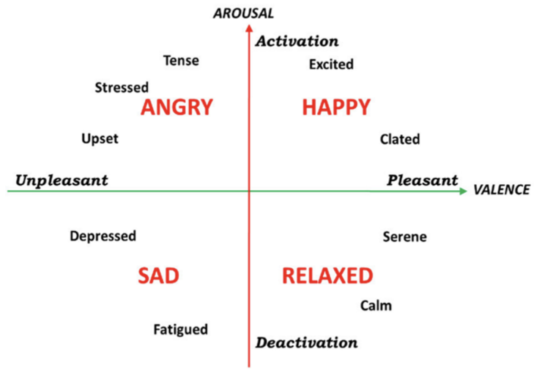

Figure 1: A Circumplex Model of Emotions (Russell, 1980).

5 Tools

- **Jupyter Notebook**: This tool will be used for data preprocessing, model training, and experimentation. It provides an interactive environment for running code and analyzing results.

- **Google Colaboratory**: A free cloud-based platform that allows users to run and collaborate on Python code. It provides a Jupyter notebook environment with pre-installed libraries, GPU support, and the ability to share and work on notebooks in real-time. Colab makes it easy to experiment and develop machine learning projects.

- **Hugging Face library**: This library offers a wide range of pre-trained transformer models, which are essential for incorporating lyrics into the recommendation system. We plan to leverage these models for text analysis, feature extraction, and building transformer-based recommendation models.

- **Genius API**: To retrieve lyrics, we utilized the Genius API. It allowed us to search and access lyrics for a vast collection of songs. By integrating the API into our project, we were able to programmatically obtain the lyrics needed for analysis and recommendation.

- **Pandas and Matplotlib**: These popular Python libraries are useful for working with data and visualizing results. Matplotlib enables us to create various types of plots and charts to visualize patterns and insights from your data. Pandas provides efficient data structures and data manipulation tools for preprocessing, organizing, and analyzing data within our project.

- **NLTK (Natural Language Toolkit)**: A comprehensive Python library for natural language processing. It offers a wide range of functionalities, including tokenization, stemming, part-of-speech tagging, parsing, and named entity recognition. NLTK also provides corpora and lexical resources for linguistic analysis and supports various machine learning algorithms for text classification and sentiment analysis.

- **SciKit-Learn**: A powerful Python library for machine learning. It provides a wide range of tools and algorithms for tasks such as classification, regression, clustering, and dimensionality reduction. Scikit-learn also offers utilities for model selection, preprocessing, and evaluation, making it a comprehensive machine learning toolkit.
- **PyTorch**: A popular open-source deep learning framework that provides a flexible and efficient platform for building and training neural networks. It combines dynamic computational graphs with automatic differentiation, making it easy to experiment, prototype, and deploy machine learning models.

6 Artificial Intelligence (AI) tools Disclosure

- Did you use any AI assistance to complete this proposal? If so, please also specify what AI you used.
  - ChatGPT

If you answered yes to the above question, please complete the following as well:

- If you used a large language model to assist you, please paste *all* of the prompts that you used below. Add a separate bullet for each prompt, and specify which part of the proposal is associated with which prompt.
  - why transformers might be better suited for lyrics based music recommendation [Introduction]
- **Free response:** For each section or paragraph for which you used assistance, describe your overall experience with the AI. How helpful was it? Did it just directly give you a good output, or did you have to edit it? Was its output ever obviously wrong or irrelevant? Did you use it to generate new text, check your own ideas, or rewrite text?
  - It was helpful in writing a cohesive answers to the prompted questions but it was always necessary to fact check it's results

References

Abdillah, J., Asror, I., and Wibowo, Y. (2020). Emotion classification of song lyrics using bidirectional lstm method with glove word representation weighting. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 4:723–729.

Agrawal, Y., Shanker, R. G. R., and Alluri, V. (2021). Transformer-based approach towards music emotion recognition from lyrics. In *Lecture Notes in Computer Science*, pages 167–175. Springer International Publishing.

Akella, R. and Moh, T.-S. (2019). Mood classification with lyrics and convnets. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 511–514.

Ali, S. (2006). Songs and emotions: Are lyrics and melodies equal partners? *Psychology of Music - PSYCHOL MUSIC*, 34:511–534.

An, Y., Sun, S., and Wang, S. (2017). Naive bayes classifiers for music emotion classification based on lyrics. In *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, pages 635–638.

Buffa, M., Cabrio, E., Fell, M., Gandon, F., Giboin, A., Hennequin, R., Michel, F., Pauwels, J., Pellerin, G., Tikat, M., and Winckler, M. (2021). The WASABI Dataset: Cultural, Lyrics and Audio Analysis Metadata About 2 Million Popular Commercially Released Songs. In *The Semantic Web. ESWC 2021. Lecture Notes in Computer Science, vol 12731.*, pages 515–531.

Chen, C.-W., Lamere, P., Schedl, M., and Zamani, H. (2018a). Recsys challenge 2018: Automatic music playlist continuation. In *Proceedings of the 12th ACM Conference on Recommender Systems*, RecSys '18, page 527–528, New York, NY, USA. Association for Computing Machinery.

Chen, C.-W., Lamere, P., Schedl, M., and Zamani, H. (2018b). Recsys challenge 2018: Automatic music playlist continuation. In *Proceedings of the 12th ACM Conference on Recommender Systems*, RecSys '18, page 527–528, New York, NY, USA. Association for Computing Machinery.

Edmonds, D. and Sedoc, J. (2021). Multi-emotion classification for song lyrics. In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 221–235, Online. Association for Computational Linguistics.

Hu, X. and Downie, J. S. (2010). When lyrics outperform audio for music mood classification: A feature analysis. In *International Society for Music Information Retrieval Conference*.

Hu, Y., Xiaoou, C., and Yang, D. (2009). Lyric-based song emotion detection with affective lexicon and fuzzy clustering method. pages 123–128.

Malheiro, R., Panda, R., Gomes, P., and Paiva, R. P. (2018). Emotionally-relevant features for classification and regression of music lyrics. *IEEE Transactions on Affective Computing*, 9(2):240–254.

Monti, D., Palumbo, E., Rizzo, G., Lisena, P., Troncy, R., Fell, M., and Cabrio, E. (2018). An ensemble approach of recurrent neural networks using pre-trained embeddings for playlist completion. pages 1–6.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178.

Szalai, G. (2020). Spotify reaches 130m paid subs, pandemic hits advertising, ceo sees on-demand audio gain.

Çano, E. (2017). Moodylyrics: A sentiment annotated lyrics dataset. pages 118–124.