T5-Based Automatic Song Title Generation

Jay Yoon, Morgan Chumney, Leah Nickelsburg, Courtney Dantzler

1 Literature Review

1.1 Song Title Generation The paper titled "AutGeLy: Automatic Title Generator Based on Song Lyrics Extractions" by Vallejo-Huanga, Careers and Mañay (2021) proposes a methodology for generating song titles. The paper employs a recommender tool, which involves loading an audio format of a song and subsequently connecting to a Shazam API to retrieve the lyrics. The lyrics information is then stored in a matrix, and the words are tokenized and normalized, with stop words being removed and content words being stemmed to reduce them to their root. The paper uses a Latent Dirichlet Allocation (LDA) model to generate a mix of words that express the song's topic, along with a Valence Aware Dictionary and Sentiment Reasoner (VADER) Python tool to perform sentiment analysis.

Our paper will adopt a similar approach to generating song titles, with several key differences. Firstly, we will begin with the lyrics as input, rather than the audio format of the song. Another significant modification we will make is with regards to the dataset size. While the original paper had a dataset consisting of 30 song lyrics, ours will consist of approximately 300 songs. This increase in dataset size is aimed at producing more accurate results by drawing from a larger and more diverse sample of songs.

Additionally, in the method proposed by the original paper, the occurrence of certain filler words in the lyrics led to flawed titles. Although stop words were accounted for, the approach did not consider the semantic significance of frequently occurring words that may not contribute substantially to the content of the lyrics. To address this issue, we propose to employ TF-IDF (Term Frequency - Inverse Document Frequency), assigning lower weights to words that occur frequently but have less semantic meaning.

1.2 Lyrics Processing Watanabe and Goto (2020) surveys the emerging field of lyrics information processing. Its approaches to semantic analysis, mood estimation, topic modeling and storyline modeling are particularly relevant to our study. Lyrics are often concise and obscure, so by cross-referencing them with a word dictionary that has valence and arousal values (similar to how Spotify annotates its audio data), researchers can understand the story of a song by the way it transitions through high and low points, such as the hook or the bridge.

Papazoglou and Gaizauskas (2021) found that incorporating listeners' interpretations of lyrics can significantly improve topic classification accuracy. The study acknowledges that most lyrics discuss multiple topics in the same song, and these topics can be clustered by both abstract concepts and concrete details by using top-n classification. It uses TF-IDF scores to calculate relevance to the topics and create word embeddings. Analysis of song lyrics has been used for various other purposes in NLP, such as identifying the age-appropriateness rating of the song (Maulidyani and Manurung, 2015), an approach which utilized text classification.

In the study conducted by Tata and Eugenio (2010), a program was developed to generate summaries of songs by extracting information from reviews of complete albums. Their approach involved a method of extraction and repackaging of individual song reviews from album reviews. Specifically, the reviews were divided into sub-sentences, each of which contained descriptions of a particular feature of a song. These sub-sentences were then repackaged into independent sentences

and reorganized into cohesive reviews. Although the focus of our work is on generating titles of songs rather than summaries, both involve the process of determining relevant information from a text and extracting and repackaging.

1.3 Other Related Works Related studies have also been done on keyphrase generation, generating important phrases to describe a lengthier textual input (Garg et al, 2022), though this approach used popular models in keyphrase generation tasks, such as catSeq, One2Set, and ExHiRD, as well as a Longformer Encoder-Decoder.

In "BUG-T5: Transformer-based Automatic Title Generation Method for Bug Reports", Tian et al (2022) uses a T5 architecture to generate titles of bug reports. The model is evaluated using ROUGE metrics, focusing on content overlap. This type of keyphrase generation differs from that of song titles in that bug reports only need to be labeled based on their objective content, while a song title generator will need to account for thematic information and artistic conventions.

2 Description of Methods

2.1 Model Architecture

The T5 (Text-to-Text Transfer Transformer) architecture uses a transformer encoder-decoder that consists of a multi-layered encoder that encodes the input sequence, and a multi-layered decoder that generates the output sequence. The encoder and decoder are connected via a cross-attention mechanism, which allows the model to effectively capture the relationships between input and output sequences.

The T5 model can be seen as a type of sequence-to-sequence (Seq2Seq) model. One of the unique features of the T5 transformer architecture is that it is trained in a text-to-text manner, meaning that the model is trained to generate output sequences from input sequences in a generalizable way. This means that the same architecture can be fine-tuned for a variety of NLP tasks, such as text summarization, question answering, and language translation.

2.2 Implementation Details

The Transformers library by Hugging Face is an open-source library that provides state-of-the-art pre-trained models for a variety of NLP tasks.

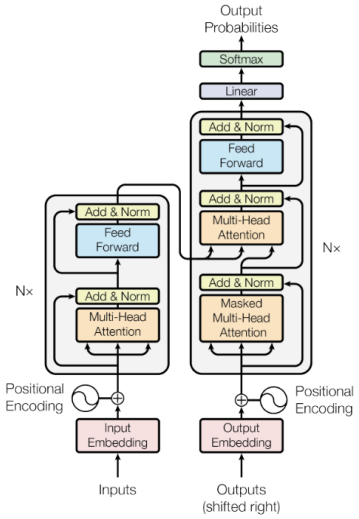

Figure 1: Encoder-Decoder Architecture

The library includes a range of transformer-based architectures, including BERT, GPT-2, and T5.

We adopt the T5 Model in the Transformers library which is pre-trained on a large corpus of text data using a "masked language modeling" task, where the model is trained to predict missing words in a text sequence. This pre-training process enables the T5 Model to learn a general understanding of language structure and context, which can then be fine-tuned for our specific task of song title generation.

2.3 Dataset

Song Lyrics Dataset obtained from Kaggle (Shah, 2018) contains over 6,000 song lyrics requested from Genius API (www.genius.com). The dataset covers a wide range of music genres, including rock, pop, country, hip-hop. In addition to this dataset, the Python client LyricsGenius allows us to scrape lyrics on our own, which we can then leverage with the Spotify API to access additional song attributes.

In our research, we exclude song lyrics whose word length is less than 100 because the summaries of such short lyrics are more likely to be incomplete and insufficient to represent the overall theme of the song. Therefore, it is assumed that songs with such few words do not require title generation. Song lyrics whose number of words exceeded 512 tokens were removed, as the T5 model has a maximum processing capacity of up to

512 tokens (Raffel et al., 2019). To further address memory constraints, we perform random sampling to reduce the size of the dataset.

Following this, we have 300 song lyrics of varied lengths. This preprocessed dataset will be used to train and evaluate the performance of our T5-based song title generation model.

2.4 Evaluation Metrics

In our study, we use the same ROUGE-N (N=1,2) and ROUGE-L metric as a measure of content overlap, with the original title as the gold-standard text. In addition to the N-gram overlap evaluation, we calculate the cosine similarity between the generated and original titles to account for semantic relatedness.

References

D. Vallejo-Huanga, E. Carrera and J. Mañay. 2021. AuTGeLy: Automatic Title Generator based on Song Lyrics Extractions. In 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), pages 1588-1593, Pasadena, CA. doi: 10.1109/ICMLA52953.2021.00254.

A.Vaswani et al. 2017. Attention Is All You Need. Computing Research Repository, arXiv: 1706.03762. Version 5.

Xinyi Tian, Jingkun Wu, Guang Yang, 2022. BUG-T5: Transformer-based Automatic Title Generation Method for Bug Reports. 3rd International Conference on Big Data, Artificial Intelligence and Software Engineering, Guangzhou, China. CUER Workshop Proceedings.

K. Garg, J. Chowdhury and C. Caragea. 2022. Keyphrase Generation Beyond the Boundaries of Title and Abstract. Computing Research Repository, arXiv: 2112.06776. Version 2.

C. Raffel et al. 2019. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. Computing Research Repository, arXiv: 1910.10683. Version 3.

Varvara Papazoglou and Robert Gaizauskas. 2021. Using Listeners' Interpretations in Topic Classification of Song Lyrics. In Proceedings of the 2nd Workshop on NLP for Music and Spoken Audio (NLP4MusA), pages 22–26, Online. Association for Computational Linguistics.

Kento Watanabe and Masataka Goto. 2020. Lyrics Information Processing: Analysis, Generation, and Applications. In Proceedings of the 1st Workshop on NLP for Music and Audio (NLP4MusA), pages 6–12, Online. Association for Computational Linguistics.

Swati Tata and Barbara Di Eugenio. 2010. Generating Fine-Grained Reviews of Songs from Album Reviews. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pages 1376–1385, Uppsala, Sweden. Association for Computational Linguistics.

Anggi Maulidyani and Ruli Manurung. 2015. Automatic Identification of Age-Appropriate Ratings of Song Lyrics. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), pages 583–587, Beijing, China. Association for Computational Linguistics.

Gaetano Rossiello, Pierpaolo Basile, and Giovanni Semeraro. 2017. Centroid-based Text Summarization through Compositionality of Word Embeddings. In Proceedings of the MultiLing 2017 Workshop on Summarization and Summary Evaluation Across Source Types and Genres, pages 12–21, Valencia, Spain. Association for Computational Linguistics.

S. Roy. "Training a T5 Transformer Model: Generating Titles from ArXiv Paper's Abstracts using Transformers", 10 November 2020, https://shivanandroy.com/transformers-generating-arxiv-papers-title-from-abstracts. Accessed 10 April 2023.

D. Shah. "Song Lyrics Dataset", 2021. https://www.kaggle.com/datasets/deepshah16/song-lyrics-dataset. Accessed 10 April 2023.