AI-Driven Song Lyrics Generation

Zeyad Alaa, Omar ElMotazBellah

May 18, 2024

1 Introduction

The field of natural language processing (NLP) has witnessed remarkable advancements in recent years, particularly in the domain of text generation. One fascinating application within this realm is the generation of song lyrics, a task that blends creativity with technical innovation. In this project, we aim to develop a generation model capable of producing song lyrics based on artist name and a starting word, leveraging a dataset of 57,650 Spotify songs.

2 Motivation

The motivation behind this project stems from the intersection of art and technology, where we seek to harness the power of machine learning to emulate the creative process of songwriting. By generating lyrics that resonate with the style and themes associated with specific artists and genres, we aim to contribute to the evolution of AI-driven creative endeavors.

Furthermore, the project offers an opportunity to explore the capabilities of neural networks in text generation tasks, ranging from traditional models to state-of-the-art architectures. Through a systematic approach encompassing data analysis, model development, and evaluation, we strive to deepen our understanding of NLP techniques and their applicability in generating coherent and engaging textual content.

3 Literature Review

3.1 Introduction to Lyric generation

Lyric generation, a pivotal aspect of songwriting, involves crafting text that conveys emotions, stories, and messages that resonate with the audience, all while aligning rhythmically and thematically with the music. Historically, this task has relied heavily on human creativity and linguistic expertise. However, with AI, the approach to lyric generation has evolved. AI models are now capable of processing vast datasets of existing lyrics to learn and mimic various lyrical styles, themes, and structures. This not only speeds up the creative process but also provides artists with a unique set of tools to enhance their creative output.

The importance of leveraging AI in these creative domains lies in its ability to extend the boundaries of artistic expression. By handling the more structured aspects of creative work, AI allows artists to focus on the core elements of innovation and personal expression. Moreover, AI-driven systems can suggest novel combinations of lyrics, rhythms, and sounds, encouraging artists to explore new artistic territories.

In the subsequent sections, we will delve deeper into the specific applications of AI in lyric generation, exploring historical perspectives, technological advancements, and the challenges that come with these innovations. We will examine how deep learning models, particularly in the context of lyric generation, have transformed the landscape of musical creativity, offering new opportunities and tools for artists to express themselves.

3.2 Historical Perspective and Evolution of Lyric Generation

The journey of lyric generation through artificial intelligence spans several decades, with each era bringing forth advancements that have progressively shaped the field. Initially, the process of generating text, including song lyrics, was based on simpler algorithmic approaches that lacked the depth and contextual awareness needed for crafting compelling and coherent lyrical content.

3.2.1 Early Methods

In the early stages, lyric generation relied heavily on rule-based systems and simple probabilistic models. These methods included techniques like Markov chains, which could generate text based on the probability of a word following a sequence of words. While these models were capable of producing grammatically correct sequences, they often lacked depth, meaning, and artistic flair, as they could not capture the subtleties of human emotion or the complex structures typical of song lyrics.

3.2.2 Transition to Machine Learning

As machine learning began to mature, so did its application in lyric generation. The shift towards more sophisticated models was marked by the use of Hidden Markov Models (HMMs) and later, more refined versions of Recurrent Neural Networks (RNNs). These models brought a significant improvement over their predecessors by better capturing dependencies in textual data, allowing for more coherent and contextually relevant outputs. However, the generated lyrics often still struggled with maintaining long-term coherence and thematic consistency.

3.2.3 Breakthrough with Deep Learning

The real breakthrough came with the advent of deep learning, particularly through the adoption of Long Short-Term Memory (LSTM) networks and, subsequently, more complex architectures like the Transformer models. These technologies introduced the ability to process and generate text with a consideration for longer narrative structures, making them ideally suited for tasks like lyric generation. LSTMs, for example, were adept at remembering and utilizing past information (like previously mentioned themes or characters) over long sequences, a crucial ability for maintaining thematic integrity in lyrics.

Throughout these developments, the key challenge has been to create models that not only produce textually and grammatically accurate lyrics but also capture the emotional depth and artistic nuances that resonate with listeners. As we have transitioned from simple probabilistic models to advanced neural networks, the ability of AI to serve as a genuine partner in the creative process of lyric writing has become increasingly realized.

In the following section, we will delve into the specific role that deep learning has played in revolutionizing lyric generation, focusing on how these models are applied and the unique capabilities they bring to the music industry.

3.3 Role of Deep Learning in Lyric Generation

Deep learning has substantially revolutionized the domain of lyric generation by equipping artists with sophisticated tools that extend beyond simple text automation to provide deeply nuanced and context-aware lyrical content. The application of deep learning technologies such as Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) network models have improved both the efficiency and quality of generated lyrics.

3.3.1 Recurrent Neural Networks (RNNs)

RNNs are pivotal in processing sequential data due to their architecture, which allows information to persist via hidden states. This capability is crucial for tasks like lyric generation, where coherence and continuity in textual data are essential. However, traditional RNNs often falter with long-term dependencies, making them less effective for longer sequences where contextual understanding is vital.

3.3.2 Long Short-Term Memory Networks (LSTMs)

To overcome the limitations of traditional RNNs, LSTMs incorporate gates that control the flow of information, significantly enhancing the model's ability to manage data over long sequences. This is particularly useful in lyric generation, where thematic and narrative consistency across verses is crucial. For instance, the paper on "Deep Learning in Musical Lyric Generation" demonstrates how LSTMs adeptly handle genre-specific linguistic nuances, adapting to varied musical styles with considerable success [GLM20].

3.4 Dataset and Features Important for Lyric Generation

The effectiveness of deep learning models in lyric generation significantly depends on the quality and characteristics of the datasets used for training, as well as the features extracted and utilized during the model training process. This section explores the types of datasets typically employed, the importance of feature engineering, and how these elements influence the performance and applicability of lyric generation models.

3.4.1 Types of Datasets

For lyric generation, datasets typically consist of a large collection of song lyrics from various artists, genres, and eras. These datasets are instrumental in training models to understand and replicate diverse lyrical styles and thematic elements. For example, in the study on "Lyrics Generator Using RNN," a dataset comprising romantic English song lyrics was used to train a model specifically tailored for generating romantic lyrics, highlighting the importance of dataset specificity in training purpose-driven models [SPK+23].

Additionally, some projects might integrate metadata such as artist names, genres, and even audio features of the songs to enrich the models' contextual awareness. This integration allows the models to generate lyrics that are not only stylistically and thematically consistent but also tailored to specific musical styles or artist characteristics, as demonstrated in the "Deep Learning in Musical Lyric Generation" paper [GLM20].

3.4.2 Feature Engineering

Feature engineering plays a pivotal role in enhancing model performance by providing models with accessible and relevant information from raw data. In lyric generation, key features often include:

- **Rhyme and Structure**: The ability to recognize and predict rhyme schemes and structural elements like verses, choruses, and bridges is crucial. Models trained on features that capture these aspects can generate more coherent and aesthetically pleasing lyrics.
- **Thematic Elements**: Features that help identify and replicate themes commonly found in lyrics, such as love, sorrow, or celebration, are essential for producing contextually relevant content.
- **Linguistic Style**: Capturing the unique linguistic style of different artists or genres, such as the use of specific colloquialisms or the frequency of certain poetic devices, helps in creating lyrics that feel authentic to the target style.
- **Word Embeddings**: These are used to transform textual data into numerical form that deep learning models can process, preserving semantic relationships between words. This is crucial for models like LSTMs, which rely on understanding contextual nuances in text.

3.4.3 Challenges in Dataset and Feature Engineering

Creating and curating datasets for lyric generation poses several challenges. Lyrics must be accurately tagged and categorized, which can be labor-intensive. Additionally, handling the diversity of lyrical content across different cultures and languages would add complexity. The choice of features also significantly impacts the model's output, requiring careful consideration and tuning to balance between overfitting on specific lyrical styles and maintaining enough generality to be creatively useful.

In summary, the quality of training data and the strategic engineering of features are fundamental to developing effective lyric generation models. These elements ensure that the models are not only technically proficient but also creatively inspiring, capable of producing lyrics that resonate with listeners and reflect the intended artistic vision. The subsequent section will delve into the specific challenges and limitations faced in lyric generation, further exploring how these issues can be addressed to enhance model effectiveness and applicability in real-world settings.

3.5 Challenges and Limitations in AI-Driven Lyric Generation

While AI-driven lyric generation offers promising capabilities, it also faces a range of challenges and limitations that affect the practicality and acceptance of the technology in the creative process. Understanding these challenges is crucial for further refining the models and ensuring their effectiveness in real-world applications.

3.5.1 Maintaining Artistic Authenticity

One of the foremost challenges is maintaining artistic authenticity. While AI models can generate lyrics that are grammatically correct and stylistically similar to human-written lyrics, ensuring that these lyrics genuinely reflect the artist's unique voice and style is complex. There is a risk of generating generic or repetitive content that lacks the emotional depth or originality that listeners expect from their favorite artists.

3.5.2 Contextual and Emotional Relevance

Lyrics are not just about words; they convey emotions and stories that resonate with the audience. AI models often struggle to fully capture the emotional nuances that are critical for impactful songwriting. Additionally, aligning lyrics contextually with the music, such as matching lyrical mood to musical tempo or genre, presents another layer of complexity that is not always successfully managed by current models.

3.5.3 Handling Diversity and Cultural Nuances

The diversity of language, cultural references, and idiomatic expressions poses a significant challenge. AI models trained on datasets from specific linguistic or cultural backgrounds might not perform well when generating lyrics for songs intended for different demographic or cultural audiences. This limitation can lead to lyrics that feel out of place or insensitive, undermining the listener's experience.

3.5.4 Scalability and Data Limitations

Scalability is another challenge, as the quality of AI-generated lyrics heavily depends on the volume and diversity of the training data. Collecting and processing large datasets of song lyrics that are diverse enough to train more universally capable models is a resource-intensive endeavor. Moreover, copyright issues often restrict the availability of such data, limiting the scope of training and the potential creativity of the models.

3.5.5 Technological Limitations

From a technical perspective, current models like LSTMs and Transformers, while powerful, still have limitations in handling long-term dependencies and complex narrative structures over extended sequences of text. This can lead to inconsistencies in longer verses or choruses where maintaining thematic or narrative continuity is essential.

3.5.6 Ethical Considerations

Finally, ethical considerations also play a crucial role. As AI begins to play a larger part in creative processes, questions about authorship, copyright, and the moral implications of using AI in art creation

become increasingly significant. Ensuring that AI is used responsibly, respecting the originality and rights of human artists, is a key concern for the industry.

In conclusion, while AI-driven lyric generation technology holds great potential, addressing these challenges is essential for its successful integration into the music industry.

3.6 Case Studies and Applications

The integration of AI into lyric generation has manifested in various innovative applications, demonstrating the versatility and transformative potential of this technology in the music industry. This section will discuss several case studies that illustrate the practical uses and impacts of AI-driven lyric generation.

3.6.1 Case Study 1: Genre-Specific Lyric Generation

As discussed in the "Deep Learning in Musical Lyric Generation" paper, AI has been employed to generate lyrics that cater to specific musical genres [GLM20]. This application involves training models on large datasets containing genre-specific lyrics, which enable the generation of content that not only aligns with the stylistic and thematic conventions of the genre but also resonates with the target audience. This targeted approach allows producers and artists to craft songs that meet market demands while maintaining creative integrity.

3.6.2 Case Study 2: Rap Lyric Generation

The "Rap Lyric Generator" study showcases an application where LSTM networks were utilized to create complex rap lyrics, emphasizing rhyme and rhythm which are essential elements of the genre [NS09]. This AI application helps artists come up with innovative lyrics that maintain the lyrical flow essential to rap music, enhancing creativity under the constraints of traditional rap structures.

3.6.3 Case Study 3: Collaborative Songwriting

In "Lyrics Generator Using RNN," AI tools were used to assist in the songwriting process, particularly helping songwriters overcome creative blocks by providing lyric suggestions based on initial inputs from the artists [SPK+23]. This collaborative tool demonstrates how AI can act as a co-creator, enhancing the creative process by introducing new lyrical ideas that songwriters might not consider on their own.

3.6.4 Case Study 4: Customization for Non-English Lyrics

The paper "AI-Lyricist: Generating Music and Vocabulary Constrained Lyrics" explores the use of AI to generate lyrics in languages other than English, specifically focusing on the unique challenges posed by linguistic and cultural nuances [MWKL21]. This study highlights the adaptability of AI systems to accommodate different linguistic structures and idiomatic expressions, providing tools for global music markets and helping artists reach international audiences with culturally resonant content.

3.6.5 Case Study 5: Emotional Tone Matching

In "The Accuracy of Recurrent Neural Networks for Lyric Generation," AI's capability to match the emotional tone of lyrics with the underlying music was examined [God18]. This case study demonstrates how AI models can be trained to understand and generate lyrics that not only fit the musical style but also enhance the emotional expression intended by the music, creating a more cohesive and impactful listener experience.

3.6.6 Case Study 6: Integrating Melody and Lyrics

The paper "A Melody-conditioned Lyrics Language Model" explores an advanced use of AI where lyrics are generated in response to specific melodies [WMF+18]. This study showcases the potential for AI to work in concert with compositional elements, generating lyrics that are not only stylistically

appropriate but also melodically congruent. This integration ensures that the lyrics enhance the musical composition, providing a seamless blend of melody and words.

3.6.7 Insights and Forward-Looking Perspectives

These case studies underscore the multifaceted applications of AI in lyric generation, from enhancing traditional songwriting processes to facilitating educational experiences. Each application not only demonstrates the practical benefits of AI but also highlights the ongoing evolution of creative processes in the music industry facilitated by technological advancements.

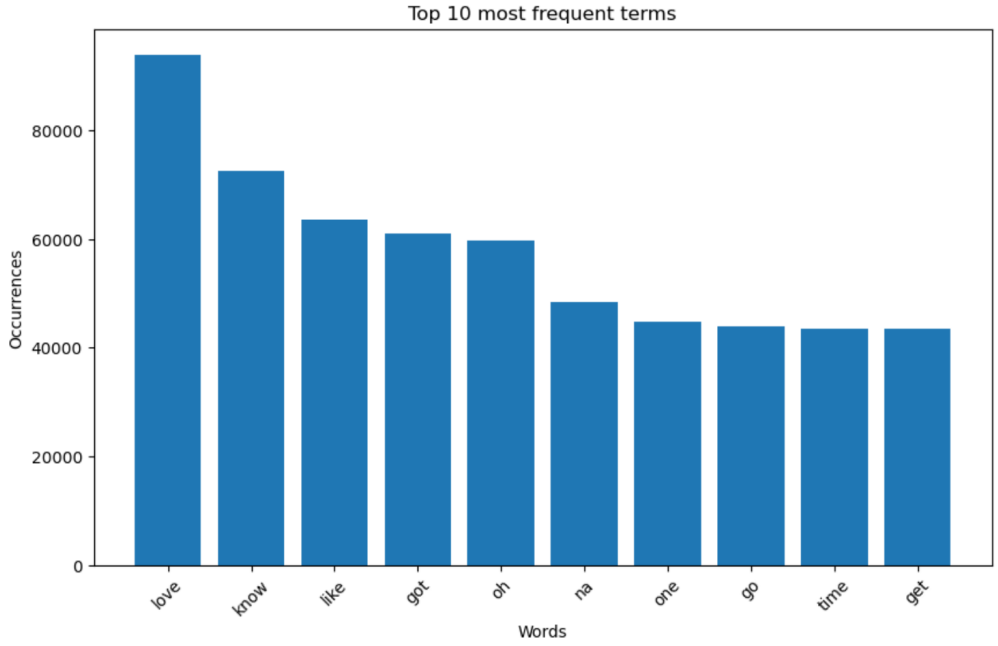
As AI technologies continue to mature, their integration into creative workflows is expected to deepen, offering even more sophisticated tools that could further revolutionize the way music is created and performed. The insights gained from current applications provide a valuable foundation for future innovations, potentially leading to new artistic expressions and collaborations between humans and machines in the realm of music creation.

4 Data Analysis and Insights

In this section, we present the results of our data analysis on the Spotify songs dataset, focusing on the most frequent terms in song lyrics and the distribution of lyric lengths. Additionally, we discuss the limitations encountered during the analysis process.

4.1 Most Frequent Terms in Song Lyrics

Upon analyzing the dataset, we extracted the most frequent terms found in song lyrics. By tokenizing the lyrics and counting the occurrences of each word, we identified the following top terms: **Frequent**



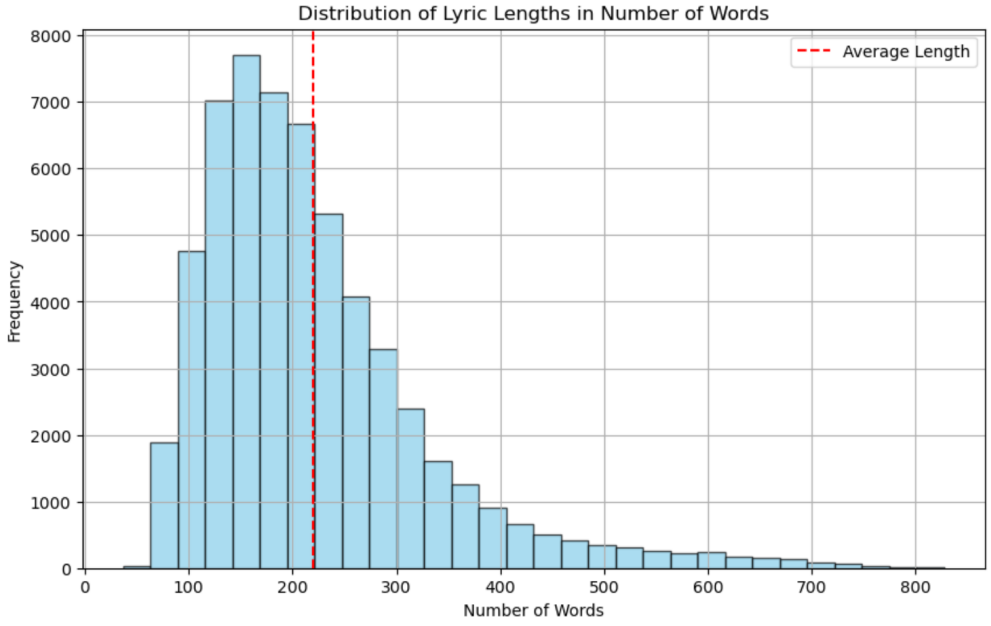
Terms:

- **love**: 93,884 occurrences
- **know**: 72,493 occurrences
- **like**: 63,567 occurrences
- **got**: 61,045 occurrences
- **oh**: 59,719 occurrences

- **na**: 48,458 occurrences
- **one**: 44,699 occurrences
- **go**: 43,826 occurrences
- **time**: 43,530 occurrences
- **get**: 43,443 occurrences

These frequent terms provide insights into the prevalent themes and motifs present in the song lyrics. Common themes such as love, relationships, and emotions are prominently featured, reflecting the universal appeal of these topics across various musical genres.

4.2 Distribution of Lyric Lengths



We further examined the distribution of lyric lengths within the dataset. By calculating the length of each lyric in terms of the number of words, we observed the following distribution:

- The average lyric length is approximately 220 words.
- The majority of song lyrics range between 100 to 300 words in length.
- There is a right-skewed distribution, with a small number of songs having significantly longer lyrics exceeding 500 words.

4.3 Limitations

Despite the valuable insights gained from our data analysis, several limitations should be acknowledged:

4.3.1 Quality of Lyrics

The quality and accuracy of song lyrics may vary, as they are sourced from diverse sources such as user submissions and online databases. This variability could affect the reliability of our analysis results.

4.3.2 Tokenization Issues

The tokenization process may not accurately capture certain linguistic nuances, such as slang, abbreviations, or irregular word forms, potentially leading to misinterpretations or underrepresentation of specific terms.

4.3.3 Sample Bias

The dataset may exhibit sample bias towards certain artists, genres, or time periods, which could skew the analysis results and limit the generalizability of our findings to the broader music landscape.

4.3.4 Lyric Length Variability

While we analyzed the distribution of lyric lengths, it's essential to recognize that song lengths can vary significantly based on factors such as musical genre, song structure, and artistic intent. Therefore, the observed distribution may not fully capture the diversity of lyrical content across different types of songs.

5 Methodology: Model Architecture

In this project, we employed a Long Short-Term Memory (LSTM) network, which is particularly suitable for tasks involving sequences, such as lyric generation. LSTM networks are a type of Recurrent Neural Network (RNN) that can learn order dependence in sequence prediction problems. This is crucial for text generation, where the context provided by previously generated words significantly influences the next word in the sequence.

Our model architecture combines two main inputs: lyric text and artist information. The text input passes through an Embedding layer followed by LSTM layers. This structure helps the model understand and generate contextually relevant text based on the input sequence. The artist input is processed through a separate embedding layer to integrate knowledge about specific artists' styles into the generation process. These two pathways converge in a Concatenate layer, combining features from both text and artist embeddings, which then feeds into a Dense layer with a softmax activation function to predict the next word in the sequence.

6 Training and Evaluation Results

The training process utilized a dataset of 57,650 song lyrics, which was first shuffled and then reduced for manageability. During preprocessing, we expanded simple repetitive structures and tokenized the lyrics to convert text into a format suitable for model training. The model was trained using categorical crossentropy as the loss function and RMSprop as the optimizer, with an initial learning rate of 0.01. Training was augmented with callbacks like Early Stopping to prevent overfitting and ensure optimal model performance.

7 Discussion of Findings and Conclusions

Our LSTM-based model demonstrates significant potential in generating stylistically and thematically consistent song lyrics. By incorporating artist-specific embeddings, the model can tailor the generated lyrics to reflect the unique styles and thematic elements associated with different artists, as seen in the generated example lyrics for "The Beatles."

However, while the model performs well in generating grammatically coherent and stylistically relevant lyrics, challenges remain in ensuring that the generated content maintains long-term thematic consistency and emotional depth. Future work could explore more complex architectures, such as integrating Transformer models, which might capture longer dependencies and subtle nuances in the text more effectively.

In conclusion, this project highlights the capabilities and challenges of using LSTM networks for artistic tasks such as lyric generation. It opens avenues for further research into more sophisticated models that could offer deeper insights and more refined control over the creative aspects of automated lyric generation.

References

[GLM20] Harrison Gill, Daniel Lee, and Nick Marwell. Deep learning in musical lyric generation: an lstm-based approach. *The Yale Undergraduate Research Journal*, 1(1):1, 2020.

[God18] Josue Espinosa Godinez. The accuracy of recurrent neural networks for lyric generation. In *Proceedings of ACM Conference*, volume 17, 2018.

[MWKL21] Xichu Ma, Ye Wang, Min-Yen Kan, and Wee Sun Lee. Ai-lyricist: Generating music and vocabulary constrained lyrics. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 1002–1011, 2021.

[NS09] Hieu Nguyen and Brian Sa. Rap lyric generator. *New York, USA*, pages 1–3, 2009.

[SPK+23] Adarsh Singh, Maharana Pratap, Abhishek Kashyap, Bhavishya Arya, Ayush Sharma, and Anand Dwivedi. Lyrics generator using rnn. 2023.

[WMF+18] Kento Watanabe, Yuichiroh Matsubayashi, Satoru Fukayama, Masataka Goto, Kentaro Inui, and Tomoyasu Nakano. A melody-conditioned lyrics language model. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 163–172, 2018.