

Generate Ballad Lyrics Using Artificial Intelligence : A Step Closer Towards Augmenting Human Creativity

Divya Singh

Walmart Labs
Divya.Singh@walmartlabs.com

Ipsita Mohanty

Walmart Labs
Ipsita.Mohanty@walmartlabs.com

ABSTRACT

There have been various approaches to augment Computational Creativity through Artificial Intelligence in the field of music. Musical creations, written by humans across diverse genres often center around consistent stories. Hence, the substantial challenge in any automated lyrics generation, particularly in the ballad genre, is addressing the most often missing element i.e contextual coherency.

MABLE (MexicA's BaLlad machinE) is an Artificial Intelligence-driven lyrics generation system that tries to overcome the above mentioned challenges in Natural Language Processing, by generating contextual coherent lyrics using machine learning models like Markov models and sentiment analysis etc.

Nevertheless, machines have always found it difficult to comprehend and evaluate the quality of output for creative domains like music. MABLE takes the human-in-loop approach to evaluate and help improve its performance over the time. This talk details the approach and architecture of this system.

AUDIENCE

This is a Beginner level talk that is targeted towards students, industry experts and artists desiring to enter the field of AI and Computational Creativity. This talk provides a strong architectural base and thought process to develop a similar and simple lyrics generator system.

INTRODUCTION

Over the past few years, AI has unlocked a box of countless possibilities in the creative realm where humans have continued to be at their best. Artists have been teaching machines about the process with the desired upshots. No wonder there is not only a need to redefine art through AI but also to specify some parameters that can measure the quality of these artworks.

Google's PoemPortraits is one such example where AI has shown tremendous abilities in generating poetry. PoemPortraits is trained via Deep Learning Neural Networks and it's collaborating feature with other people around the world generates consistently growing collective poems.

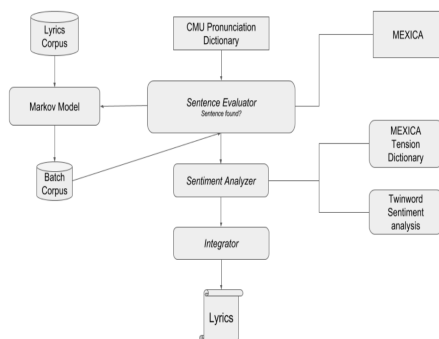
Similarly, our architecture comprises two systems - MABLE and MEXICA that interact with each other collaboratively to generate creative art forms. MABLE relies on MEXICA for its input to generate lyrics. And finally, the human-in-loop component ensures its improved quality of art generation.

MEXICA is a plot generation system based on the Engagement-Reflection model for creative writing. It produces novel coherent stories about the Mexicas, speaking indigenous people of Mexico City. In MEXICA, each story is defined as a sequence of actions. Each action

consists of a set of preconditions and postconditions in terms of emotional links and tensions between characters. For example, the precondition of the action “the hunter murdered the princess” might be the emotional link “the hunter hates the princess”, the post-condition of the action “the eagle knight awarded the princess” might be the emotional link “the princess is very grateful to the eagle knight.” This emotional information is used by MABLE to extract sentiments present in the storyline.

Plot generation by MEXICA is followed by a sentence evaluator component of MABLE that generates a list of sentences based on a corpus of lyrics using Markov Model. So, for each of the lines in the story, MABLE finds the following promising line thus playing a part of a collaborator. The quality of the line is determined by factors such as rhyme quality, number of syllables, rhyming schemes, and rhyme type. Then, the sentiment analysis and integration of sentences come into the picture.

We have three main components in MABLE’s architecture: Sentence Evaluator, Sentiment Analyzer, and Integrator. Let us discuss each of the components in detail.



Sentence Evaluator uses a second-order Markov model to generate a list of candidate lines. The model is trained on a corpus consisting of lyrics of 29 love songs sung by various artists. Scrapy is used to crawl 70-80s, pop songs available on azlyrics.com which is then followed by data cleaning. For each line generated by MEXICA, the Sentence Evaluator gives a batch of phrases. Seeing not getting good results in one loop, the model is repeatedly called until a set of at least 50 high-quality candidate lines

are formed. The quality of those lines is calculated on the basis of the rhyming score and the number of syllables. To get the rhyming quality, NLTK is used to create a pool of words that rhyme with the last word in the storyline, taken from MEXICA, to various degrees. CMU Pronunciation Dictionary provides multiple phonetic transcriptions of words. For each of the pronunciations, a tuple list of candidate words and their rhyme qualities are calculated. To further optimize the rhyme quality number, the degree of phonetic similarities is taken into account, such as end rhymes, slant rhymes, and assonance rhymes. And then, the syllable count provided by CMUdict is also taken into consideration to choose the good ones.

Sentiment Analyzer takes the output from Sentence Evaluator and connects those emotionally to the story. Twinword API is used to extract sentiment scores for each word in the sentence. The values are then averaged to get the emotional quality of each sentence and sentences are then clustered into positive, negative, and neutral groups. MEXICA also provides us with emotional information contained in the story in the form of tensions such as love competition or clash in emotions. Then, it is made sure that sentiments of lines from Sentence Analyzer and MEXICA fall in the same range.

MEXICA generates stories in the third person, most of the lyrics on which Markov model is trained are in the first person. **Integrator** neutralizes the point of view by changing from first to third-person point of view.

At the end of our compositions, there came a question of how to evaluate our art-work. To evaluate the quality of compositions, we involved humans and showed people the lyrics generated by MABLE, other AI systems as well as human-made lyrics. For each of the artworks, we asked participants whether they are made by machine or an artist and asked them to evaluate them on the basis of some already set evaluation parameters. The results are explained in detail in the next section.

Few of the previously related works included approaches that were mainly limited to template-based and statistical methods and hardly used the artificial-intelligence driven end-to-end solutions. Additionally, none of the previous works ensured that the systems evaluated the quality of their artworks to improve over time.

Examples:

- (1) Tra-la lyrics 2.0 combines two systems, PoeTryMe and Tra-La-Lyrics where user-provided words are fed into a sentence generator that uses templates and a semantic network to form sentences.
- (2) Full- face poetry extracts key phrases from newspapers which are then combined with a database of similes to create template-based lines.
- (3) SMUG (Scientific Music Generator) takes out keywords from academic papers and then fills predefined song structures.
- (4) DopeLearning uses deep neural net and SVM models to construct from phrases selected from a database of rap lyrics.

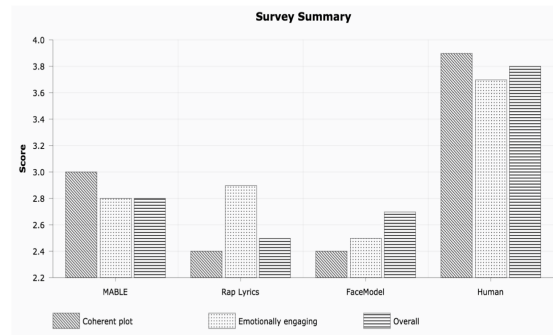
OUTCOMES/CONCLUSION

For any artwork generated by an AI system, the evaluation part is the most challenging part. A user study evaluation was performed comparing MABLE's lyrics with two other machine generated lyrics - Rap Lyrics and Facemodel along with the ones written by humans.

The survey-takers population was selected randomly from various different backgrounds. The final drawn sample size consisted of 30 students out of which 26 were undergraduate and 4 were graduate students, with an average age of 24. Among these 30 students, 26 were male and 4 were female.

Survey-takers were blind to the creators of the lyrics. They had to rate these lyrics based upon three parameters:- emotional engagement, coherence, and overall quality of lyrics. Answers were provided via the standard 5 option Likert Scale, rating quality from very good (5) to very poor (1).

The results are laid out in the form of the graph shown below.



The sample survey summary shows that out of all the machine generated lyrics, MABLE generated lyrics were rated highest for coherency and overall quality of lyrics. However, Rap Lyrics are very emotionally engaging amongst lyrics composed by all systems. Like many other Word-level models, MABLE too suffers from overall low quality of lyrics when compared against Human written ones. Comparatively saying, Character-level models have proven to be more impressive.

Our approach establishes a nuanced start towards moving the machines a bit closer to human creativity. Some of the thoughts around our future work includes,

- (1) Feeding a lyrics corpus of higher language complexity and emotion to train the models.
- (2) Close human-in-loop by feeding in the evaluation of MABLE's lyrics back into the Markov Model as a decision making feature to improve the quality of lyrics generated.
- (3) Evaluating Long Short Term Memory (LSTMs) models as a replacement/augment for Markov Models.

While our system is not perfect and has it's fair share of technical challenges, it's also undeniable that we can't completely teach machines to imitate human creativity. Nevertheless, we could always depend on machines to augment our creativity to generate even more powerful pieces of art-works in the days to come.

PARTICIPATION STATEMENT

If our paper gets accepted, we will be there to present our thoughts.



REFERENCES/BIBLIOGRAPHY

- Ballad Lyrics Generator Paper: [\[PDF\] A Ballad of the Mexicas: Automated Lyrical Narrative Writing](#)
- MEXICA creative writing: [MEXICA: A computer model of a cognitive account of creative writing](#)

BIO

Divya Singh, Software Engineer, Walmart Labs

She has been a Software Engineer at Walmart Labs since Aug. 2018. She enjoys learning something new every single day which gives an ultimate sense of achievement and satisfaction as a professional. She did research work under the supervision of Dr Maya Ackerman in the field of computational creativity during her graduate school years at San Jose State University.

Ipsita Mohanty, Software Engineer - Machine Learning, Walmart Labs

A Carnegie Mellon University graduate with 10+ years of software industry experience across diverse domains. Ipsita is a Software Engineering & Machine Learning professional, who believes that Machine Learning at scale is a new way forward. Additionally as a frontier tech hobbyist, she believes that AI and creativity amalgamation will lead to discovery of unthinkable horizons in the days to come.