CS395T Project Proposal

Lucas Kabela and Bill Yang

September 24th 2020

1 Goal Of the Project

This project aims to make progress on conditional text generation through exploring improvements in conditional lyric generation. In particular, we aim to improve on previous attempts at conditioning on artist, while most prior efforts on enforcing theme and rhyme scheme in lyric generation.

Conditioning generation on persona has seen success in improving consistency and engagement in dialog response agents (Liu et al. 2020 [[5]]) as well as improvements in text style transfer. As such, we hypothesize conditioning on artist / persona will provide improvements to the quality of lyrics generated in conditional lyric generation as well.

2 Motivation

Early models for long-text generation were able to produce syntactical and fluent text, but their output lacks direction and is not natural. Conditional text generation has recently become a popular technique for controlling long text to provide a tune-able direction of output. We however identify 2 key areas that differentiate lyric generation from other long text generation tasks: songs have more structure than summaries, stories, and other long-texts and lyrics follow rhyme schemes.

Prior approaches to lyric and text generation have explored incorporating attributes to steer generation for the issues mentioned. However, these approaches lack consistency with respect to artists. There are many factors that differentiate the style of artist and this can have a downstream effect on the structure of the lyrics generated, and rhyme scheme (for instance, an artist like MF Doom has much denser rhyme scheme than Lil Peep). As such, we seek to evaluate and understand the influence of artist and artist personas as conditioning in lyric generation, which prior works have overlooked.

3 Related Work

First, we aim to explore works focused on lyric generation. These texts are a mixture of unconditional and conditional generation attempts (Potash et al. 2015 [10]), (Malmi et al. 2016 [6]). (Nikolov et al. 2020 [8]), (Wu et al. 2018 [15]). Most of these approaches are transformer/seq2seq based, but there has been some work in variational auto encoders in this field as well.

For our specific problem, (Vechtomova et al. 2018 [14]) attempts to perform genre conditioning, which is a broader version of our work. Rhyming is explored in (Hopkins et al. 2017 [3]).

There is a also a recent paper on conditioned poem generation which we hope to use for direction in enforcing structure in our work (Yang et al. 2018 [16]).

Lastly we have a variety of papers that deal with authorship of text, personas, and style transfer: (Manjavacas et al. 2017 [7]), (Tikhonov et al. 2018a [11]), (Oliveira et al. 2015 [9]), (Tikhonov et al. 2018b [12]), (Li et al. 2016 [4]), (Fu et al. 2017 [2]), (Liu et al. 2020 [5]). We hope that exploring these tangential areas will provide much more insight into how to structure our models and data to increase the influence of the artist persona.

4 Dataset and Proposed Architecture

For our dataset we will use Genius API [1], which has a large collection of annotated songs and artist profiles. We choose to scrape Genius over existing datasets for two reasons - first, Genius contains a large catalog of music, and includes contemporary album releases as well. Thus, Genius provides the largest corpus of data available, and the api is fairly simple to use. Second, Genius provides a large amount of labels for our training data in the form of annotations and artist profiles. These labels are provided by both crowd-sourcing to fans, and from the respective artists themselves, thus making them a strong source of annotated data. We plan to use all songs in an artist's discography, with verses from featured artists skipped over during training and test time.

Our proposed architectures will largely evolve on existing methods. We will begin by using an LSTM based approach (Potash et al. 2015 [10]) as our baseline. Our other model we will use will will be a transformer based method

[1] https://docs.genius.com/

(Vaswani et al. 2017 [13]) incorporating prior efforts for lyric generation and personas (Li et al. 2016 [4]). We also seek to use an attribute model as another extension (Dathathri et al. 2020 [1]). Time permitting we may consider using the CVAE based model(Yang et al. 2018 [16]) to explore a less conventional approach.

5 Methodology and Evaluation

For our methodology, we will first run generation with these models with no artist conditioning to provide a strong baseline for our metrics under no conditioning. This will effectively recapture results for prior works under our metrics. Next, we will perform an ablation study on the artists personas - starting with just a token for artists, and expanding into an "artists profile" which will include city, record label, and the name of prior albums. This "profile" will be flexible, depending on the ease of capturing information from Genius.

For input to our models, we will use the conditioned artists information, with appropriate tags for the models to learn as seen in the persona literature, and the first line of the song. The lines generated will serve as recurrent input to the model, until a full verse is produced. We plan to use teacher forcing to steer the model appropriately.

The main things we want to measure is if our generated text seem like song lyrics, if the lyrics are unique, and if the style is similar to the given artist. Rhyme density (RD) is a common metric used in lyric evaluation to see if fluent rhyme scheme has been achieved, which captures the "lyricness" of our text. We plan to use distinct N-gram/self-BLEU matching to determine the uniqueness of the lyrics generated. However, artist vary in their repetitiveness, so (Potash et al. 2015 [10]) presents an automatic metric to measure an artist similarity score, and uniqueness of the lyrics generated which we plan to use as a stronger metric. Sample cross entropy (Tikhonov et al. 2018b [12]) can also be used to determine the similarity between texts based on their artists, and we will incorporate this metric if the automatic metric in Potash et al. 2015 is insuffiecent to capture relationships.

Since lyrics are highly subjective, we will also plan to get a small set of annotators (nonprofessional) to provide a likert scale score for our generated lyrics in terms of "goodness". We expect the deviation to be quite high for this metric. To ground this, we will also provide an A/B form of evaluation where actual lyrics from the artists are provided with our generated ones.

6 Conclusion

In conclusion, we aim to explore the effectiveness of conditioning lyric generation on artist persona. Whereas prior works in lyric generation largely overlook the impact artist can have on rhyme scheme and lyrical structure, we believe our work can provide a natural bridge between prior works on conditional generation and enforcing structure in generated data. We gather our data from Genius API, and use LSTM as a baseline model, with transformers and techniques in the literature providing an upperbound on performance. For evaluation, we will perform ablation studies as to the level of conditioning and its impact, as well as several metrics to measure artist-likeness, rhyme density, and distinctness.

References

[1] Sumanth Dathathri, Andrea Madotto, Janice Lan, Jane Hung, Eric Frank, Piero Molino, Jason Yosinski, and Rosanne Liu. Plug and play language models: A simple approach to controlled text generation, 2019.

[2] Zhenxin Fu, Xiaoye Tan, Nanyun Peng, Dongyan Zhao, and Rui Yan. Style transfer in text: Exploration and evaluation. *CoRR*, abs/1711.06861, 2017.

[3] Jack Hopkins and Douwe Kiela. Automatically generating rhythmic verse with neural networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 168–178, Vancouver, Canada, July 2017. Association for Computational Linguistics.

[4] Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. A persona-based neural conversation model. *CoRR*, abs/1603.06155, 2016.

[5] Qian Liu, Yihong Chen, Bei Chen, Jian-Guang Lou, Zixuan Chen, Bin Zhou, and Dongmei Zhang. You impress me: Dialogue generation via mutual persona perception. In *Proceedings of the 58th Annual Meeting of*

the Association for Computational Linguistics, pages 1417–1427, Online, July 2020. Association for Computational Linguistics.

[6] Eric Malmi, Pyry Takala, Hannu Toivonen, Tapani Raiko, and Aristides Gionis. Dopelearning: A computational approach to rap lyrics generation. *CoRR*, abs/1505.04771, 2015.

[7] Enrique Manjavacas, Jeroen De Gussem, Walter Daelemans, and Mike Kestemont. Assessing the stylistic properties of neurally generated text in authorship attribution. In *Proceedings of the Workshop on Stylistic Variation*, pages 116–125, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.

[8] Nikola I. Nikolov, Eric Malmi, Curtis G. Northcutt, and Loreto Parisi. Conditional rap lyrics generation with denoising autoencoders, 2020.

[9] Hugo Gonçalo Oliveira. Tra-la-lyrics 2.0: Automatic generation of song lyrics on a semantic domain. *J. Artif. Gen. Intell.*, 6(1):87–110, 2015.

[10] Peter Potash, Alexey Romanov, and Anna Rumshisky. GhostWriter: Using an LSTM for automatic rap lyric generation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1919–1924, Lisbon, Portugal, September 2015. Association for Computational Linguistics.

[11] Aleksey Tikhonov and Ivan Yamshchikov. Sounds wilde. phonetically extended embeddings for author-stylized poetry generation. In *Proceedings of the Fifteenth Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 117–124, Brussels, Belgium, October 2018. Association for Computational Linguistics.

[12] Alexey Tikhonov and Ivan P. Yamshchikov. Guess who? multilingual approach for the automated generation of author-stylized poetry. *CoRR*, abs/1807.07147, 2018.

[13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *CoRR*, abs/1706.03762, 2017.

[14] Olga Vechtomova, Hareesh Bahuleyan, Amirpasha Ghabussi, and Vineet John. Generating lyrics with variational autoencoder and multi-modal artist embeddings. *CoRR*, abs/1812.08318, 2018.

[15] Xing Wu, Zhikang Du, Y. Guo, and H. Fujita. Hierarchical attention based long short-term memory for chinese lyric generation. *Applied Intelligence*, 49:44–52, 2018.

[16] Xiaopeng Yang, Xiaowen Lin, Shunda Suo, and Ming Li. Generating thematic chinese poetry with conditional variational autoencoder. *CoRR*, abs/1711.07632, 2017.