Generating Artist Styled Song Lyrics Using a Bi-Directional Long Short-Term Memory Neural Network

John Angle
UCCS
jangle@uccs.edu

Tyler Wengert
UCCS
twengert@uccs.edu

Abstract

Since the early days of Natural Language Processing (NLP), poetry and text generation has been used as an exercise in creating more human-like text with a computer. Many modern approaches achieve this by providing a neural network with a huge corpus of human written texts for it to draw from. A distinct type of poetry, song lyrics are one of the most heavily artist stylized forms of writing, and the average person can easily recognize the lyrics of some of the most popular contemporary musicians. In this paper we provide an end of the term wrap-up on our Bi-Directional LSTM Chorus Generator, analyze the results from the model, and outline improvements to the model that would potentially achieve state of the art results and make it a candidate for publication. The final LSTM array outputs very recognizable choruses, and it is clear after testing that the model just needs to receive some more targeted data to improve its weaknesses; the addition of multiple pre-processing models that can analyze and append further information to the corpus would likely shore up these gaps in the models knowledge.

1 Introduction

While poetry generation has been a topic of research since the late 80's, stylized output received little attention until the mid 90's when a group of meteorologists used NLP text generation techniques to output scientifically formatted weather reports (Goldberg et al., 1994). The field saw rapid acceleration in the 2010's when machine learning, artificial intelligence, and NLP were brought together, and the corpus methodology for giving a system content and form was introduced (Toivanen et al., 2012). One thing that many modern text or poetry generation systems have in common is a lack of any real sense of author voice, style, or authenticity. When it comes to the applications of computer generated text, a realistic tone and sense of author can drastically improve the quality of a service offered. Customer service or emergency response systems, faithfulness to a sense of voice across translations, and general language studies are a few worthy motivations.

The keys to producing high quality text with a better sense of voice are to provide the predictive model with an appropriate corpus that models the desired author's style, as well providing it with as much information about the relationships in the text as possible. Our models are built on corpora whose content and size was specifically tailored to this task, and so gives us a strong base with which to work. In addition, we tie together multiple well-explored NLP tools to produce a final output that achieves more than any of these models could on their own, as well as lay out steps to integrate more tools and knowledge into the package. In the final model, these elements worked well together to produce coherent choruses, and we expect that this synergy between elements will continue to produce extremely high quality text that is consistent across even more aspects of style as more tools are implemented in the chain.

2 Related Works

Since our array of Bi-Directional LSTM neural networks will be the driver of our system, a strong language model needed to be developed to produce as high quality text as possible. Among the many different types of neural network layers, LSTMs have distinguished themselves as exceedingly useful in the text processing and Natural Language Processing. An extension of the classic LSTM, Bi-Directional LSTMs feed the training content forwards and backwards through the layer, thereby capturing more relationships than the standard LSTM. In their 2017 paper, Xie et al. achieves outstanding results generating Shakespearean Sonnets using a word-based LSTM system. The model uses a single LSTM layer, fed by a custom embedding layer. This word based model was able to generate original, coherent sonnets from a corpus of Shakespeare's works. Tested against other RNN and CNN frameworks, the standard Gated LSTM performed better across categories of 'Coherence', 'Poeticness', and 'Meter/Rhyme'. Tikhonov and Yamshchikov (2018) used a similar model soon after, extending the approach by using a dual LSTM setup, and adding sentence embeddings, document embeddings, and a soft-max layer after the two LSTMs. Their results in generating multilingual author-stylized poetry are state of the art. Our model will be a word-based one, rather than character-based, following results from (Sutskever et al., 2011) and (Choi et al., 2016).

Part of speech tagging, (POS tagging) aims to label each word in a corpus with its corresponding English Language part of speech. This particular field of NLP gave researchers problems for many years, due to the variable relationship between a word in its part of speech. The same word can take on multiple different parts of speech depending on the context. For many years, statistical models were favored, the Hidden Markov Model (Blunsom, 2004) being one of the most popular and successful. In recent years, neural networks have advanced POS tagging greatly. (Kim, 2014) (2014) achieved state of the art results using a Convolutional Neural Network (CNN) setup. Recent models have made adjustments and improvements to Kim's model to advance the state of the art.

Sentiment analysis is a field of NLP that focuses on gleaning meaning from the context of words. In particular, can we figure out whether a word is positive or negative, and how greatly so. Salehin et al. (2020a) tested an LSTM-based Recurrent Neural Network on a corpus of Bengali Facebook posts, and had a high level of success (at least 72.86% accuracy) in scoring words by their sentiment. Salehin et al. (2020b) also used an LSTM-based system to produce the most recent state of the art results.

Potash et al. (2016) propose a method for generated rhymes. This metric gives researchers a baseline with which they can evaluate their generated rhymes. To achieve these standards, (Malmi et al., 2016) used a phonetic based Seq-2-Seq model to detect a variety of rhyme types, including perfect rhyme, consonance, and assonance. They use a line by line generation method to output rhyming poetry/song lyrics. Their tool has been released as an online tool called 'DeepBeat' and is constantly being analyzed and improved.

3 Implementation

This program was built using Python 3.7.6 64-bit using the IDLE IDE. The program uses the NumPy package (Oliphant, 2006) for handling the data arrays, and NLTK (Loper and Bird, 2002) for preprocessing of the corpora.

The main driver of the system is the array of Bi-Directional LSTM neural networks that predicts the generated text. While its output is aided and formatted by other components of the system, its predictions are the driving force for the generated choruses. The LSTM models are trained on small, artist specific corpora, and given a large knowledge base through (Mikolov et al., 2013) word vectors. The various components of the system are described in further detail below.

3.1 LSTM Array

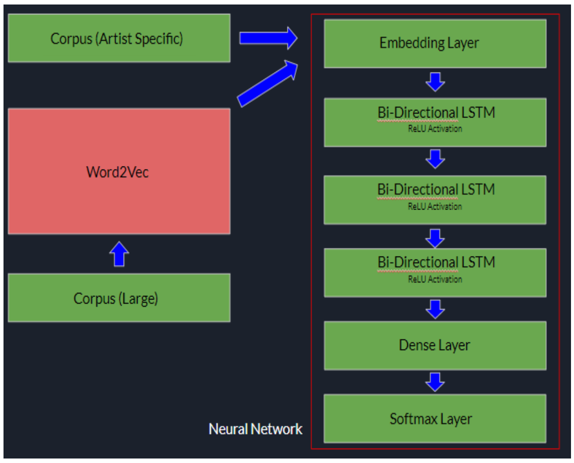

Figure 1: Design of the Bi-Directional LSTM network

The neural networks are implemented using the Keras platform from (Chollet et al., 2015) and the TensorFlow platform from (Abadi et al., 2015). The program also makes use of NumPY for working with the one-hot encoding used for the model, as well as pickle for saving and loading the model.

Each Bi-Directional LSTM is dedicated to a specific artist, and is trained on a corpora of the artists 50 most popular songs. Specifics about the data set can be found in 3.2.

The corpus pre-processing included removing non-ASCII characters, extraneous escape characters, and tokenizing the text. Punctuation characters, including the comma, question mark, and exclamation point, were broken down as a unique token rather than being included with the word proceeding it. Contractions (don't, can't, wouldn't, etc.) were treated as a single word rather than as a base and terminal pair as in some papers. We don't ever want the word 'don' or the word "'t" to appear on their own, so keeping them as a single token will eliminate that chance.

The model consists of 10 layers. After preprocessing, the corpus is passed into an embedding layer. Here, the corpus dictionary is matched with its Word2Vec vector, from the model described in 3.3. This layer has its 'trainable' parameter set to false; we want to use the pre-learned word vectors without the Embedding layer training them as it would if there were random or 0 initialized embeddings fed to it. The embedding layer feeds these embeddings to the first of the Bi-Directional LSTMs as 100 dimension vectors. The model is designed as a 3-stack of these Bi-Directional LSTM layers, each followed by a Dropout layer, with a dropout rate of 0.3. The first two Bi-Directional LSTM layers each have their 'return_sequences' parameter set to true, to allow the data to be fed into another LSTM layer; The final LSTM does not use this, as it feeds into a Dense layer with 300 dimensions. The final layer in the model, another Dense layer, has dimensions equal to the number of words in the corpus vocabulary. This dense layer uses a soft-max activation. Each network was trained for 50 epochs. The models all attained high marks in categorical accuracy, all the scores falling in a range from 77 - 88 percent. It is also noteworthy that despite each model being trained on a different data set, they all managed to achieve accuracy rates in a fairly tight range. This suggests that our model's design is well suited for the data we are using.

For the models predictions, it is important to note that the model was initialized with a random number of lines to output, as well as a random number of words in each line. The line count is limited to the range 3 - 5 (inclusive), since we want to replicate only choruses, not entire verses. To improve the coherency of the lines, it would be better for the model to decide when a line should terminate. To accomplish this, the model would be trained with the newline character ('
n') included as a vocabulary word. Then, the newline character would be a perfectly valid word that the model could predict, and it would not need to be initialized with a random number.

When having the model output its predictions, another important consideration is the temperature that you use. A temperature value below one will reduce the variety of the predictions, making the highest probability prediction more likely to be selected, whereas a temperature value above one will increase variety by increasing the likelihood that lower probability predictions will be selected. A single temperature value did not produce optimal results for all of the models equally, so some testing with each model was needed to find a temperature that worked for that particular artist. The table below details the various temperature used for each artist's model.

Queen	1.05	The Beatles	0.5
Green Day	0.5	Brad Paisley	.75
Carrie Underwood	0.5	Florida Georgia Line	1.0
Kendrick Lamar	1.2	Eminem	.9
Lil Jon	1.0	Bruno Mars	.75
Adele	1.2	Britney Spears	.33

Table 1: Temperatures for each model

3.2 Data Set

Genius, a website that advertises itself as the world's largest collection of song lyrics, gives users access to a Python based API, and it was from here that we pulled our corpora. We had two distinct types of corpora: The large, mixed genre corpus and the much smaller, artist specific ones.

The general corpus consisted of lyrics from 4 distinct genres of music: Rock, Country, Pop, and Rap. Within these genres, we selected the top 50 songs from every distinct artist on the top 20 lists from 2006 until 2019. This corpus consists of about 3.2 million words, and has about 47,000 distinct vocabulary tokens. Following from the results of (Godinez, 2018), we also built smaller corpora consisting of the most representative material from each of 12 selected artists. These 12 artists were chosen as candidates due to their distinct, recognizable personal styles. The 12 artists we will be mimicking are:

- Selected Artists: The Beatles, Queen, Green Day, Florida Georgia Line, Carrie Underwood, Brad Paisley, Kendrick Lamar, Eminem, Lil Jon, Bruno Mars, Britney Spears, Adele

Billboard has been a music industry standard for reporting top artists and songs since the 1940's (Molanphy, 2013). Their charts are the basis for the general corpus artist selections.

- Rock: Panic! At The Disco, Queen, Imagine Dragons, twenty one pilots, The Beatles, Portugal. The Man, Five Finger Death Punch, Linkin Park, Metallica, Coldplay, X Ambassadors, James Bay, Fall Out Boy, Hozier, WALK THE MOON, Lorde, Bastille, Green Day, Coldplay, Passenger, The Lumineers, Phillip Phillips, Mumford Sons, fun., The Black Keys, Gotye, Foster The People Foo Fighters, Florence + The Machine

- Country:Luke Combs, Dan + Shay, Kane Brown, Thomas Rhett, Florida Georgia Line, Chris Stapleton, Sam Hunt, Blake Shelton, Carrie Underwood, Zac Brown Band, Jason Aldean, Eric Church, Taylor Swift, Hunter Hayes, Lady Antebellum, Rascal Flatts, Keith Urban, Sugarland, George Strait, Tim McGraw, Kenney Chesney, Brad Paisley, Toby Keith

- Pop: Ariana Grande, Khalid, Jonas Brothers, Halsey, Dua Lipa, Camila Cabello, Maroon 5, Ed Sheeran, Bruno Mars, The Chainsmokers, Shawn Mendes, Alessia Cara, Justin Bieber, Selena Gomez, Adele, The Weeknd, Nick Jonas, Katy Perry, OneRepublic, Sam Smith, Justin Timberlake, Macklemore Ryan Lewis, Flo Rida, Rihanna, Britney Spears, Lady Gaga, Ke$ha, Jason Derulo, Taio Cruz, The Black Eyed Peas, Beyonce, Chris Brown, Leona Lewis, Jordin Sparks, Jesse McCartney

- Rap: Cardi B, Meek Mill, DaBaby, Travis Scott, Lil Baby, Drake, Post Malone, Migos, Kendrick Lamar, Future, Big Sean, Desiigner, DJ Khaled, Fetty Wap, Nicki Minaj, Rae Sremmurd, Kid Ink, JAY-Z, Iggy Azalea, YG, ScHoolboy Q, Logic, Lil Wayne, J. Cole, 2 Chainz, Wiz Khalifa, Ludacris, Young Money, B.o.B, T.I., Kanye West, Plies, T-Pain, Unk, Bow Wow, Mims, E40, Lil Jon, Eminem

There were a handful of artists that appeared in the top charts of more than one genre; their name only appears in the most representative list here, and their songs were only included once (see Taylor Swift is both a Pop and a Country artist; Post Malone is both a Pop and a Rap artist; Mumford & Sons is both a Country and a Rock artist). Since these four genres were selected on the hypothesis that they will be particularly distinct, we will avoid using any artists that appear in multiple genres as a target for stylized output and analysis. They are included in the general corpus because we feel that their music and lyrics will still be highly representative of the genres we are modelling, and they will add value to the word vectors they are used to train.

3.3 Word2Vec Vectors

The data fed into the Word2Vec model was preprocessed in the same manner as the data fed into the LSTM models. Again, when tokenized, punctuation symbols were treated as distinct tokens from the words that proceeded them. The word vectors were trained using the Continuous Bag-of-Words (CBOW) method, in which the model is fed context words in a specified window size and asked to predict the middle word. The vectors were defined to have 300 dimensions, and were trained with a window size of 10 over 300 iterations.

While we wanted the final predictor model to output words over a tighter artist vocabulary, we also didn't want to limit the amount of knowledge that the model would learn about the structure and flow of lyrics. To accomplish both these goals, we decided that the transfer learning provided from the Word2Vec vectors would allow the model to capture the information we wanted it to from the large corpus. It also helps prevent the model from spitting out s verbatim from the original corpus by introducing a little more randomness.

3.4 Lexical Diversity in the Corpora and the Output

We inspected our selected musical genres for lexical diversity and vocabulary. The data was preprocessed in the same manner as in s 3.1 and 3.3.

Rock	Country	Pop	Rap
.0562	.0480	.0432	.0538

Table 2: Main corpus lexical diversity values

Rock and rap have a notably higher diversity than country and pop, which may be strong enough to distinguish the pairs of genres. The next step is a closer inspection of vocabulary, whether certain keywords are specific to one genre or other. To show this, not only must many occurrences be represented but occurrences should be shown across multiple artists within the genre and not in other genres.

4 Continuing Work

4.1 Part of Speech Tagging

One of the most noticeable flaws in the output from the Chorus Generator is a lack of proper English grammar. While it is certainly not required or even always desired for song lyrics to adhere strictly to grammar rules, but there are certainly occasions when the model predicts a word that doesn't make much sense in terms of its 'grammar context' and it throws the natural flow of a phrase off. Upon examination of many of the output samples, it seems that a large majority of these instances are caused by words that take on different meanings depending on the part of speech they take on in a given context. Providing part of speech tagging (POS tagging) information to the predictive models, both the Word2Vec and LSTM, these instances could be reduced or eliminated.

The implementation of this would not require any modification to the current system, only the addition of the POS tagging element prior to the training of the model. To achieve high accuracy in POS tagging, a Convolutional Neural Network (CNN) approach based on (Kim, 2014). The corpus would first be fed to this POS tagging network, where each word is associated with its respective tag. When feeding data to the predictive LSTM model for training, the word and its POS tag will be taken together as a single token. This will have the effect of increasing the unique vocabulary size some, but should reduce out of context words that break the flow of a chorus.

4.2 Sentiment Analysis

A key aspect of musical style is the sentiment and tone that the artist conveys throughout their lyrics. It is important for conveying meaning and keeping the musical flow. In a single chorus, the sentiment expressed and the tone will usually stay consistent throughout, and that is what we are looking to capture.

To achieve this, we will implement a system similar to that of (Salehin et al., 2020a), who used the model to classify the sentiment in Facebook posts. They created an LSTM based classifier to rank to words as either positive or negative. Words will be scored on a scale from -1 to 1, where -1 indicates strongly negative and 1 indicates strongly positive. Similarly to the POS tagging above, the corpus will be run through this sentiment analyzer before being passed to the predictive model for training. The sentiment score would be appended to each word and passed to the predictive model as a single token. As before, this will cause the vocabulary size to grow, but will allow the model to differentiate words by a greater number of parameters, and allowing for predictions that rely not solely on the word, but also on more detailed relational information about it and the words around it.

4.3 Rhyme Schemes

Recognizable lyrics also require a mixture of rhyming. Some rap artists display their style by packing as many rhymes into a line as possible, while most pop music employs a repeating, multi-line rhyme pattern. This can make a chorus more memorable, more poetic, and is another dimension of style any songwriter must consider.

Recognition of rhymes can be seen as part of a predictive labelling or translation task, referred to as grapheme-to-phoneme (G2P). G2P seeks to label words with their pronunciations, and simply matching these labels can then produce rhyming. While Rao and others detailed an effective LSTM approach to G2P models, our hope is to use pre-trained G2P models available in Python to produce phoneme labels, such as g2p-seq2seq (Rao et al., 2015).

To produce rhyming structures for generated lyrics, we will focus only on line endings. We label rhymed line endings with A, B, C and so on, and label same word endings with X1, X2, X3. Consider this chorus from Queen:

We are the champions, my friends
And we'll keep on fighting 'til the end
We are the champions
We are the champions
No time for losers
'Cause we are the champions

In terms of rhymed line endings, there are three distinct phonemes. We would label the rhymes, in order, as A, A, B, B, C, B. We also have three lines with a same word ending, and we signify this structure as $X1, X2, X3, X3, X4, X3$. With these templates, we can attempt to conform generated lyrics to artist specific rhyming and sameness. This could be approached as a matter of post-processing, or we might include phoneme tagging similar to POS tagging for training the predictive LSTM model.

5 Evaluation

To evaluate the poems, we had a group of 5 test subjects complete a survey analyzing various aspects of the poems. The test group was composed of our roommates and family. Unfortunately, due to Corona-virus situation, it was difficult to reach many participants. The survey tested the choruses on two levels: whether or not the chorus was well written, and whether or not it successfully emulated its author's style.

3 categories were defined in which a poem must score well to be considered a well written chorus: readability, content, and emotion. Readability concerns the structure of the chorus. Do the sentences flow correctly? Do adjacent words make sense together? And least importantly, since strict compliance to English grammar rules are not necessary for song lyrics, is it written with proper grammar? Content deals with the meaning and context of the poem. Does the chorus keep a consistent context (for example, a song references different leisure activities that occurred during the weekend) or does it mix unrelated contexts? Does the chorus give more meaning as a whole than the words would individually? Finally, emotion deals with the reader's response to the generated chorus. Did the poem elicit emotions from the reader? How intense were those emotions? Participants will rate each chorus in these three categories on a scale from 1 - 5, with 1 being poor representation of the category's qualities, and 5 being equivalent to human written text. The overall score for a poem is the average of its scores in the three categories.

In addition to the chorus being well-written and making some sense, we want to know if it successfully emulated its author's style. Participants were presented with one of the generated choruses, and a selection of four artists. The participants were asked to pick the artist that they thought most likely to have written the lyrics, and then rate how confident they were in that choice on a scale from 1-5, where 1 indicated no confidence whatsoever, the selection may as well have been a guess, and 5 indicated absolute certainty that the chosen artist was correct.

Participants were given a selection of 12 generated choruses, one from each selected artist, in a random order. No participant was given the same generated choruses as any other participant. Before any of the test questions, participants were given a written explanation of the various evaluation categories. The format of the test questions can be found in Appendix B.

5.1 Results

The choruses that the current predictive model has put out are of quite high quality. The model did an excellent job of imitating author style, with varying results across genres, and was also able to maintain sufficient coherency to be readable and convey some meaning. In this section, we analyze some specific examples and key data points, for additional chorus samples and a full list of the participant results, refer to Appendices A and C

In Section 3.4, we discuss the lexical diversity between the genres. Looking at the results of the model, it seems that the original lexical diversity of a genre has some bearing on the quality of the produced text. Pop noticeably had the lowest lexical diversity of the four genres, and it really showed in the outputs. When generating choruses with the Pop artists, they were the most likely to get stuck in repetitive loops. 'Oh oh oh oh...', 'yeah yeah yeah...', and 'hey you hey you' were phrases that were very commonly predicted for all three of the Pop artists we analyzed. The results of the survey reflect this as well, where 2 out of the 3 Pop artists scored lower than average for selection confidence. The correct artist was also not identified quite as frequently for Pop artists as other artists. On the other end of the spectrum, Rap had one of the higher lexical diversities, and as a result there was a correspondingly high lexical diversity shown in the outputs. Not only was was Rap as a genre very distinguishable from the other three genres, but the artists within the genre were also incredibly distinguishable from each other.

The models were able to pick up some interesting patterns and latent features of the lyrics. One literary device that some of the artists utilize in their lyrics is consonance. Lines with lots of consonance often flow well and have a steady cadence. In particular, Brad Paisley and Eminem are both known for their use of consonance. Many lines output from these two artists' models exhibited consonance to varying degrees, like this one imitating Brad Paisley, "cassette in a pontiac town snack wears a ballcap boots and jeans and doesn't break some glass". This line exhibits strong consonance on the hard 'a' sound in pontiac, snack, cap, and glass. Many of the artists had themes that were referenced frequently, and this instilled a deeper sense of authorship into the choruses. The Beatles love to talk about love: "love love love you everybody love you", "you need me like to love you". They also frequently mention light, dark, and the juxtaposition between the two: "comes the sun and i go ", "singing in the light he up in the dark black night ", "in the dark lies beneath the ocean waves".

All the artists selected primarily create their music in English, however, some of the artists lyrics contained a small number of lines in a foreign language; in most cases, Spanish. The model handled these cases more adeptly than expected, and predicted Spanish words that made sense in context with both the English and Spanish words surrounding it. It may be interesting to extend the model to include entire foreign language corpora in future iterations.

While the model has done an excellent job at capturing the spirit of the artist, it still leaves some room for improvement in terms of its readability and grammar. None of the artists scored poorly on the writing quality metrics, but none of them scored particularly well, either. As discussed in Section 4.1, the addition of a POS tagging system during the data pre-processing stage is likely to improve this aspect of the generator.

6 Conclusion

Overall, this system was a success. We managed to chain a couple different NLP tools together to get the benefits of each of them. The generated choruses largely accomplish our goal of being attributable to an artist, although that is not to say there is no room for improvement. The Rhyme Analyzer component is one upgrade the system could receive to better capture style. The largest area for growth is in the readability and coherency of the generated texts. As we continue to improve the system, we must be careful to balance the need for grammatical structure and the need for the freedom from the strictest parts of English Language rules that lyrics require.

References

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

Phil Blunsom. 2004. Hidden markov models. *Lecture notes, August*, 15(18-19):48.

K. Choi, G. Fazekas, and M. Sandler. 2016. Text-based lstm networks for automatic music composition. *arXiv preprint arXiv:1604.05358*.

François Chollet et al. 2015. Keras. `https://keras.io`.

J.E. Godinez. 2018. The accuracy of recurrent neural networks for lyric generation. *Proceedings of ACM Conference 17*.

E. Goldberg, R. Kittredge, and N. Driedger. 1994. Using natural-language processing to produce weather forecasts. *IEEE Expert*, 10(2):45–53.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar. Association for Computational Linguistics.

Edward Loper and Steven Bird. 2002. Nltk: The natural language toolkit. In *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics - Volume 1*, ETMTNLP '02, page 63–70, USA. Association for Computational Linguistics.

Eric Malmi, Pyry Takala, Hannu Toivonen, Tapani Raiko, and Aristides Gionis. 2016. Dopelearning: A computational approach to rap lyrics generation. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 195–204.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

C. Molanphy. 2013. How the hot 100 became america's hit barometer. *NPR - The Record*.

Travis Oliphant. 2006. NumPy: A guide to NumPy. USA: Trelgol Publishing.

P. Potash, A. Romanov, and A. Rumshisky. 2016. Evaluating creative language generation: The case of rap lyric ghostwriting. *University of Massachusetts Lowell*.

K Rao, F Peng, H. Sak, and F. Beaufays. 2015. Grapheme-to-phoneme conversion using long short-term memory recurrent neural networks. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, QLD*, pages 4225–4229.

SM Samiul Salehin, Rasel Miah, and Md Saiful Islam. 2020a. A comparative sentiment analysis on bengali facebook posts. In *Proceedings of the International Conference on Computing Advancements*, pages 1–8.

SM Samiul Salehin, Rasel Miah, and Md Saiful Islam. 2020b. A comparative sentiment analysis on bengali facebook posts. In *Proceedings of the International Conference on Computing Advancements*, pages 1–8.

I. Sutskever, J. Martens, and G.E. Hinton. 2011. Generating text with recurrent neural networks. *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 1017–1024.

Alexey Tikhonov and Ivan P Yamshchikov. 2018. Guess who? multilingual approach for the automated generation of author-stylized poetry. In *2018 IEEE Spoken Language Technology Workshop (SLT)*, pages 787–794. IEEE.

J.M. Toivanen, H. Toivonen, A. Valitutti, and O. Gross. 2012. Corpus-based generation of content and form in poetry. *International Conference on Computational Creativity*, pages 175–179.

Stanley C Xie, Ruchir Rastogi, and Max Chang. 2017. Deep poetry: Word-level and character-level language models for shakespearean sonnet generation.

A Sample Generated Choruses

Sample choruses are organized by genre and artist. We intentionally kept the examples in the main body of this paper free from profanity, however, some examples in this appendix include profanity. These are the author's original words, and nothing was removed or censored, as we felt that vocabulary choice, including profanity, was integral to conveying an author's style. As a general rule of thumb, Pop is mostly clean, Country and Rock can go either way, depending on the artist, and Rap is the most profane.

A.1 Rock

A.1.1 Queen

- now dead all dead and gone
  now but one real folks are make you standing
  me down to fly away minah
  minah ooh ooh ooh ooh ooh mister void ooh if ooh

- hey hey ow ow man gotta getta me gotta get out
  but lad standing me just gotta get out no right outta here
  i got no right from the typhoon plan pray for the story tellers run you

- i hear ride my ritz till the
  end of time takeoff hey hey hey hey hey hey
  hey hey hey hey hey hey hey hey yeah hey

- will you do the fandango will be in
  your seaside take my farthing get me
  down to mars on the bough yeah yeah yeah yeah yeah

A.1.2 The Beatles

- can enjoy the show it is shown you're melting
  out in miami behind his storm he
  creeps from a noise comes
  from behind the behind them five ago

- bang bang shoot shoot shoot coca cola
  coca cola bye across the universe a shoulders
  comes the sun and
  i go carrying pictures of chairman mao mao

- love love love you everybody love you
  singing in the light he up in the
  dark black night in the dark lies beneath
  the ocean waves lies beneath the ocean waves
  are he noise we'll go out

- you need me like to love you
  singing in the light in the dark black night
  he was quizzical studied pataphysical science
  in the home sweet home

A.1.3 Green Day

- dragging my mother's love to the lights
  of blood that declare i read
  the blood so we've been victimized

- tonight i got me too well hid nowhere i don't lose
  my home to you live on my own here and bleed
  the moment has that's on my own beat
  my mind ooh ooh ooh

- all the demons without me
  exist in a bay ooh ooh ooh ooh ooh
  ooh i love you love i'm still shame the shame
  it's war jimmy and it's just dangerous ah ah

- salmonella poison oak no home i'm leaving away to
  testify and quite still vietnam but never forget
  me told my first time of hell don't care
  burning down me out of treason right now i know

A.2 Country

A.2.1 Brad Paisley

- cassette in a pontiac town snack wears a ball cap boots
  and jeans and doesn't break some glass in
  a window down the street down there there in
  the moonlight whoa got a snack

- we're heather that's true it's just another world in love me and you'll see me
  in the light of the silver linin' if he did cause you don't jump
  in the other love the first girlfriend if you feel how he's grown

- in world two two is a job interview than
  what i leave it can be shockin' be
  stuck in the temper and my first dog you wish you could see

- if be crushin' it there's a bunch of weight everytime i go in love and log in the shoes
  and that's a snack smoked joint to be
  a hundred few invisible white hoods and invisible white hoods and chevrolet his
  skies and invisible white hoods in his bottom
  white roads in his desk on

A.2.2 Carrie Underwood

- to see my face and close my face and
  close your eyes and they appeared outta known hand and
  drug my shade of joy and they find humble
  knockin' into the tone and didn't
  say out and i just let one old

- 'round when they go back go by asking the
  prayers fell to the night they possess out of his babies
  singin' singin' singin' singin' singin' singin'

- poison and on the dark newspapers floor our
  side to the side of the color of the night to the
  backseat underneath a bee town singin' singin'

- it feels our heart it's our favorite part of the story it's
  our trauma veils didn't bother
  to cry one horse fallin' and here in
  repeat on repeat on repeat

A.2.3 Florida Georgia Line

- to get me created to get your hands on you get me right before
  you wanna slip on me
  tin roof cuttin' real slick wake up and see the world and
  wake up and see wake in the world spinnin' spinnin'
  digits right aww yeah see comin' how to have
  to stand it through the years

- my 5 racing like driving a microwave line just
  embrace the fact straight up again
  as it don't can go straight
  up floatin' of shot a microwave or the sticks his luck on

- wide her lips like lighters like them bales ride a drink from a drink like us
  wake in the ride coke again grew up into me i wake up

- we travel to go alone and both got here changing
  lanes to hang in a microwave of george
  car from the bar a reflection or me again every reflection
  that confession you didn't think i've been fifty miller high life like a gypsy lot

A.2.4 Rap

A.3 Kendrick Lamar

- dot you tryna bleed him back by his knees nigga she mmm
  homie facts when that make
  some car fuck coward your drink with it homie you don't believe it like

- i'm really really real i'm really really real i'm really really real i'm really
  really real i'm really really real i'm really
  really real 'cause i'm real i wanna
  see your brother on you i'ma be alright alright huh i'm 'bout to be peak

- like traffic on it i
  wanna be with this earth nigga go on cloud 9
  since my cousin seen two two times two man two cell three power him two
  what's five two demarcus shambles

- filled with stash collected temptation is explosives with demons come and hide my feelings it's
  give every music works to you on the sense of me tell two karma and
  live by it i'm eatin' the feds guess we hit sure you
  see his fragrance it in the denim of god snatch you on

A.4 Eminem

- static and i hate to have what destiny you cussin' i guess
  i'ma go and fuck themselves with
  a fuckin' colonic mockingbird call me cujo
  stairlift stairlift beware of me

- boca mejor maricon little puto and all of you culo joyner joyner
  low yeah yeah yeah y'all seen profit 'til the shit 'til i spit
  the hook motion off go fly back

- is impossible i'll let it outside up with a pen pace i'm dope with a
  thot afraid of clickbait that'll die intent at anything you feel
  stupid he choppy did it though fuck you damn you could hurt me at
  a pen no no y'all does
  beautiful me to do you with

- out size
  i lay your ass sprayed like pam
  done in vain half the fuck you 45

A.5 Lil Jon

- this motherfucker then i motherfucker then shake that
  shit while you ride the d ride da d ride
  da d ride da d ride da d ride da d ride da d ride

- in my mind and a body body like a body body
  like she workin on the pole representin' that thang
  on that jock chocolate waka flocka waka flocka

- the start this start the top baby just the bass just drink
  on the back smack dat ass and run your eyes on me it's gametime boy yeah yeah yeah

- me digas nada cuando te agarre
  ey boy ai culo culo
  culo shake that ass

A.6 Bruno Mars

- i'm out here wanting to
  moon call the night is gone talking to the

moon stops and the floor
baby mirror is what you ever

- night long as i'm gone
room and let the floor it should
bells sing like ooh yeah yeah you did it this is cool coming on the
door ow it feels it's out for i ever world that i have

- you better better run you better run run run away run away baby i've
done there i i've done
of a eye before the friends before i do i never was the same don't preach two is what you

- gone room starin' at you and they gotta decide the call
the cops but it feels
don't do not give up yeah

A.7 Adele

- bring it home i miss it i miss it i will be
the truth that i adjust use to heal the truth that i use to
heal the pain two feet lea the stars lea lea the river lea

- two feet appear they will always be rain where we will be
claim you prefer where you come around in the
worst of me see me right again wait there will be blame you on the river

- and the despair of your fingertips and town 'cause you went on over
everything now they thought of me in town claim me over they started all
of the bricks of everything changed me all
of your clue mouth that town without town anymore

- of my love opened up my bed as mistakes mistakes bitterness
where we'll took my bed as across the
bed as i'm concerned the we'll breathe in the stars

A.8 Britney Spears

- let me see you dance on the
floor baby baby if i'm the
way i feel the jokes with the same closer i like you dunno

- high feels so good i don't wanna stop i go there innocent
'cause i could be acting bizarre everything to trouble bath with your face
it's haunting me what i'm saying i'm flipping me off limits

- there's only two types of people in
the world pretty girls are dirty luv like a little bit
of danja danja danja

- take it down if you think i start to run away with
me my slow mind you jokes tonight there's no froze me spinning and
she just can't stop that stealing i go

B Test Question Format

Chorus: "Generated Chorus Here"

Rate the above song chorus on a scale from 1-5 for the following categories. Refer to the beginning of the survey packet for explanation of each category:
Readability:
Content:
Emotion:

Select the artist who you feel is most likely to have written this chorus.
a. Artist1
b. Artist2
c. Artist3
d. Artist4

Rate on a scale from 1-5 how confident you are in your selection:

C Results Continued

C.1 Participant #1

Artist	R	C	E	Avg
Queen	3.5	3.0	2.9	**3.13**
The Beatles	3.7	3.2	3.3	**3.4**
Green Day	3.5	3.1	3.1	**3.23**
Brad Paisley	3.8	3.4	3.3	**3.5**
Carrie Underwood	3.3	2.8	3.0	**3.03**
Florida Georgia Line	3.0	3.3	3.0	**3.1**
Kendrick Lamar	4.0	3.5	3.5	**3.66**
Eminem	3.9	3.6	3.3	**3.6**
Lil Jon	4.0	3.6	3.5	**3.7**
Bruno Mars	3.3	3.1	2.9	**3.2**
Adele	3.3	3.3	3.0	**3.2**
Britney Spears	3.0	2.7	3.4	**3.03**

Artist	Correct?	Conf
Queen	0	**2**
The Beatles	1	**4**
Green Day	1	**3**
Brad Paisley	0	**1**
Carrie Underwood	1	**2**
Florida Georgia Line	0	**1**
Kendrick Lamar	1	**4**
Eminem	1	**5**
Lil Jon	1	**4**
Bruno Mars	1	**4**
Adele	0	**3**
Britney Spears	0	**2**

C.2 Participant #2

Artist	R	C	E	Avg
Queen	3.4	3.2	3.2	**3.26**
The Beatles	3.4	3.2	3.5	**3.36**
Green Day	3.3	3.1	3.4	**3.26**
Brad Paisley	3.8	3.6	3.6	**3.66**
Carrie Underwood	3.4	3.3	3.3	**3.33**
Florida Georgia Line	3.5	3.3	3.4	**3.4**
Kendrick Lamar	4.2	4.0	3.8	**4**
Eminem	4.1	3.7	4.1	**3.96**
Lil Jon	3.8	4.1	4.1	**4**
Bruno Mars	3.3	3.2	3.1	**3.2**
Adele	3.2	3.3	3.3	**3.26**
Britney Spears	3.0	3.1	3.4	**3.16**

Artist	Correct?	Conf
Queen	0	**2**
The Beatles	1	**5**
Green Day	0	**3**
Brad Paisley	0	**2**
Carrie Underwood	1	**4**
Florida Georgia Line	1	**4**
Kendrick Lamar	0	**3**
Eminem	1	**4**
Lil Jon	1	**3**
Bruno Mars	0	**1**
Adele	0	**2**
Britney Spears	1	**4**

C.3 Participant #3

Artist	R	C	E	Avg
Queen	3.3	3.5	3.2	**3.33**
The Beatles	3.4	3.4	3.3	**3.36**
Green Day	3.2	3.3	3.4	**3.3**
Brad Paisley	3.6	3.6	3.6	**3.6**
Carrie Underwood	3.3	3.4	3.2	**3.3**
Florida Georgia Line	3.5	3.2	3.3	**3.33**
Kendrick Lamar	3.7	3.8	4.0	**3.83**
Eminem	3.9	3.9	3.8	**3.86**
Lil Jon	3.8	4.0	3.8	**3.86**
Bruno Mars	3.4	3.4	3.4	**3.4**
Adele	3.3	3.1	3.4	**3.26**
Britney Spears	3.1	3.0	3.0	**3.03**

Artist	Correct?	Conf
Queen	1	**4**
The Beatles	1	**4**
Green Day	1	**3**
Brad Paisley	1	**4**
Carrie Underwood	0	**2**
Florida Georgia Line	0	**1**
Kendrick Lamar	0	**2**
Eminem	1	**3**
Lil Jon	0	**3**
Bruno Mars	1	**4**
Adele	0	**2**
Britney Spears	1	**4**

C.4 Participant #4

Artist	R	C	E	Avg
Queen	3.2	3.2	3.1	**3.16**
The Beatles	3.3	3.4	3.4	**3.36**
Green Day	3.1	3.2	3.4	**3.23**
Brad Paisley	3.7	3.5	3.4	**3.53**
Carrie Underwood	3.3	3.2	3.2	**3.23**
Florida Georgia Line	3.2	3.3	3.3	**3.26**
Kendrick Lamar	3.7	3.5	3.7	**3.63**
Eminem	3.8	3.4	3.6	**3.6**
Lil Jon	3.8	3.8	3.5	**3.7**
Bruno Mars	3.3	3.3	3.2	**3.26**
Adele	3.1	3.1	3.4	**3.2**
Britney Spears	3.0	2.9	3.3	**3.06**

Artist	Correct?	Conf
Queen	1	**5**
The Beatles	1	**5**
Green Day	1	**4**
Brad Paisley	0	**1**
Carrie Underwood	1	**3**
Florida Georgia Line	0	**1**
Kendrick Lamar	1	**3**
Eminem	1	**2**
Lil Jon	0	**2**
Bruno Mars	1	**3**
Adele	0	**3**
Britney Spears	1	**4**

C.5 Participant #5

Artist	R	C	E	Avg
Queen	3.3	3.2	3.3	**3.26**
The Beatles	3.3	3.4	3.6	**3.43**
Green Day	3.1	3.4	3.5	**3.33**
Brad Paisley	3.7	3.6	3.6	**3.63**
Carrie Underwood	3.3	3.4	3.2	**3.3**
Florida Georgia Line	3.3	3.3	3.3	**3.3**
Kendrick Lamar	3.8	3.6	3.9	**3.76**
Eminem	3.8	3.6	3.8	**3.73**
Lil Jon	3.8	3.8	3.4	**3.66**
Bruno Mars	3.3	3.2	3.5	**3.33**
Adele	3.3	3.3	3.4	**3.33**
Britney Spears	3.2	3.1	3.5	**3.26**

Artist	Correct?	Conf
Queen	1	**5**
The Beatles	1	**5**
Green Day	0	**2**
Brad Paisley	1	**4**
Carrie Underwood	1	**5**
Florida Georgia Line	1	**4**
Kendrick Lamar	0	**1**
Eminem	1	**2**
Lil Jon	0	**1**
Bruno Mars	1	**4**
Adele	1	**3**
Britney Spears	1	**5**