RNN-Stone Cowboy

"I told him it was not the perfect country & western song
Because he hadn't said anything at all about mama
Or trains, or trucks, or prison, or getting' drunk"
- David Allan Coe, "You Never Even Called Me by My Name"

Introduction

In recent years, AI researchers have made great strides in applying neural networks to sequence based, variable-size input problems. Models like RNNs, LSTM, GRUs, and transformers have been successfully applied to a wide variety of data types, including audio, video, and most notably, text.[1][2] In text specifically, RNNs and transformers can translate between languages, generate document summaries, and compose articles, poetry, and music.[3][4][5][6][7]

The purpose of this project is to generate novel lyrics in the country genre using the Long Short-Term Memory (LSTM) deep learning architecture, a leading architecture for sequence-based problems. While LSTM has been overtaken by more complex transformer models (e.g. BERT, GPT-2) in the last several years, LSTM can be implemented from scratch, providing customizability and learning opportunity while still retaining strong performance. I experimented with two models. The first was a basic approach using one layer of LSTM, no dropout, and random sequence selection for batch creation.[8] The second approach utilized two layers of LSTM, and ordered batch creation.[9]

In addition to these two models, I expanded on the traditional text generation process by parameterizing the generation process with topic modeling and by introducing stochasticity to expand the vocabulary during generation using word embeddings.

Data Sources and Preparation

Data Acquisition: The corpus used for model training and evaluation consists of the majority of songs featured in the Billboard Hot Country Charts from 1959 to April 2022. Song title, artist, and rank from each weekly chart were pulled from the Billboard website via web-scraping. From here, I aggregated chart data into a table of unique songs by title and artist and pulled the lyrics to each unique song using the lyricsgenius

---

[1] Gimeno, P., Viñals, I., Ortega, A. et al. Multiclass audio segmentation based on recurrent neural networks for broadcast domain data. J AUDIO SPEECH MUSIC PROC. 2020, 5 (2020). https://doi.org/10.1186/s13636-020-00172-6
[2] Liu, F., Chen, Z. & Wang, J. Video image target monitoring based on RNN-LSTM. Multimed Tools Appl 78, 4527–4544 (2019). https://doi.org/10.1007/s11042-018-6058-6
[3] Sutskever, I., Vinyals, O. & Le, Q. V. (2014). Sequence to sequence learning with neural networks. Advances in neural information processing systems (p./pp. 3104--3112), .
[4] Jingqiang Chen, Hai Zhuge. Extractive summarization of documents with images based on multi-modal RNN. Future Generation Computer Systems, Volume 99, (2019). https://doi.org/10.1016/j.future.2019.04.045
[5] Porplenko, D. (2020). Generation of sport news articles from match text commentary.
[6] Ghazvininejad, M., Shi, X., Choi, Y., & Knight, K. (2016, November). Generating topical poetry. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 1183-1191).
[7] Kento Watanabe, Yuichiroh Matsubayashi, Satoru Fukayama, Masataka Goto, Kentaro Inui, and Tomoyasu Nakano. 2018. 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, New Orleans, Louisiana. Association for Computational Linguistics.
[8] This approach was largely drawn from the architecture used in the third homework of the Deep Learning course in the UVA SDS Masters Program (DS 6050).
[9] This approach was inspired by the architecture used in Chapter 16 of Aurélien Géron's Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition

package developed by genius.com. Several pre-processing steps were then necessary due to data quality issues and model architecture needs.

Data Cleansing: Lyricsgenius does not always match songs to lyrics correctly, and even when it does the results occasionally contain errors. Songs were deemed to be "valid" if they had appropriate word, line, and section structure (under 20 sections per song, under 50 lines per section, under 1000 words per line). These were conservative estimates, but helped to catch the vast majority of incorrectly pulled lyrics (e.g. novels, TV episode transcripts, etc). Punctuation was removed and lyrics were converted to lowercase. From here, songs were vectorized at the word level, using the symbols *{"<l>", "<s>", "<e>"}* to denote the end of lines, sections, and songs, respectively. Words with invalid characters (e.g. emojis, Chinese characters) were replaced with the "unk" token.

Vocabulary Reduction: Since my model was designed to predict words rather than characters (see "Methodology" section), it was necessary to reduce the vocabulary size to create a manageable final dense layer. I reduced my vocabulary to the top 1000 tokens, and replaced all valid out-of-vocab tokens (e.g. no invalid characters) with the closest in-vocabulary token in the word embedding space as determined by the GLOVE values from the Wikipedia dataset at the 25-dimension level. The "unk" character was again used when none of the 100 closest words in the embedding space were in-vocab.

Train-Test Split: In both cases, 90% of the data was used for training and 10% for validation. This splitting was done at the year-song level: 90% of songs from each year were used for training and 10% for validation. This was done to ensure that each song's lyrics stayed intact, and that the data was not biased towards any particular time period.

Batching: Two approaches to batching were implemented. The basic approach used sequences of 750 words for each observation, while the advanced approach used a sequence length of 200. In both cases, batches of size 16 were utilized. These decisions came as the result of hyperparameter tuning. In the basic approach, batches were chosen at random from the input space at each training pass. In the advanced approach, epochs were used, and sequences were kept in corpus order for each batch index. This was done to ensure stateful continuity throughout the training steps.

Methodology

Model Architecture: Models were implemented in the keras module in tensorflow. The basic approach featured an embedding space of size 256, one layer of LSTM, and a dense layer of 1001 to match the vocabulary size (1000 words plus "unk"). The advanced approach followed the same architecture, this time using two LSTM layers to improve its capacity for complexity. Both approaches used stateful training, which retains cell state from one batch to the next during training. Both models used the keras default glorot uniform recurrent initializer and sigmoid recurrent activation.

Training: Both models used Adam as an optimizer, and an exponentially decayed learning rate beginning with $1x10^{-3}$. Training was conducted until either training performance reached a plateau or I ran out of time.

Generation: Much of this project's unique contribution is manifested in the generation procedure. In addition to a simple in-vocab "next word" selection procedure, this implementation added parameterization and stochasticity. Parameterization came in the form of topic modeling. I ran Latent Dirichlet Allocation (LDA) on the corpus using scikit-learn's LDA module, producing a 40-topic language model.[10] From these, I explored the top word in terms of information gain (Equation 1) in each topic, and manually created labels

[10] Blei et al. Latent Dirichlet Allocation. Journal of Machine Learning Research 3 (2003) 993-1022

for these (e.g. *{"Tennessee", "Texas", "southern", "south"}* is tagged as "the South"). LDA produces a topic-word matrix of probabilities for each word in the vocabulary. To parameterize topics for lyric generation, the output probabilities for each word were multiplied by their topic probabilities to create a final distribution from which the chosen token was sampled (Equation 2).

$$Information\ Gain = \frac{p(w_i|topic_z)}{\sum_z p(w_i|topic_z)} \quad (1)$$

$$p(token) = softmax_{tokens}\left(p(token|topic) \times p(token_{LSTM})\right) \quad (2)$$

In addition to parameterization, I intended to stochastically increase the output vocabulary using GLOVE word embeddings accessed using the genism package. For each "chosen" word, the generation function pulled the embedding vector, using this as the mean of a multivariate Gaussian from which the true output word was generated. A diagonal covariance matrix with a value of 0.2 (another tuned hyperparameter) was used for this. To maintain functionality of the limited-vocab network, the "chosen" word was passed through to the next generation step rather than the true output word.

Results

Model Performance: Both models performed poorly. Testing loss vastly exceeded training loss, a clear sign of overfitting, and the lyrics generated by both models amounted to little more than nonsense (see examples in Appendix). In the case of the "advanced" model, testing loss continued to rise dramatically even as training loss dropped (Figure 2). Interestingly, an early iteration of the simple model produced similar training loss to the final (I had not yet added validation assessment at this time) and produced much structurally sound and somewhat grammatically accurate lyrics. Unfortunately, this model was lost due to issues with saving in Google Collab, but the moderate success of this early iteration supports the choice of word-level LSTM to approach this problem.

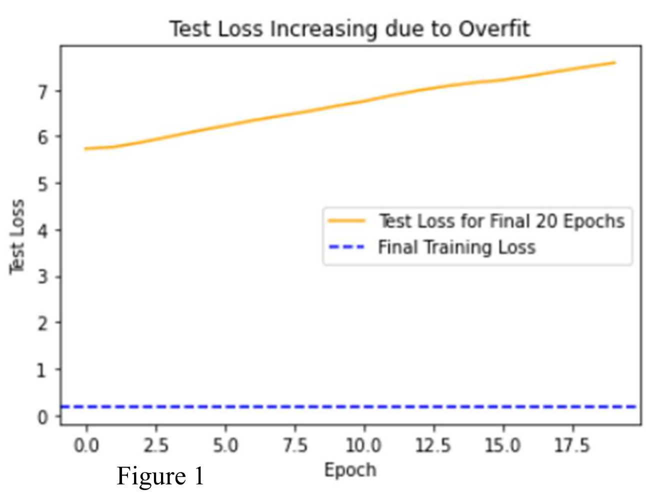

Figure 1

Parameterization: When the lyric generation process functions properly (as in the lost implementation), the topic model adjustment to the base generation procedure works well – the grammatical performance is similar and the words chosen fit the theme. When the lyric generation process works poorly, the results are similar to random sampling using the topic model probabilities.

Vocabulary Enhancement with Word Embeddings: It is difficult to assess the performance of the vocabulary enhancement step because this was not yet implemented when base generation was producing meaningful results. It does, however, seem that this process is prone to major issues, as many of the words chosen through the stochastic process were clearly off-model, including nonsense words and foreign characters.

Discussion and Future Work

I suspect that the failure of the advanced model can be explained by important differences between my implementation and the prototype textbook implementation on which it was based. That version trained and produced results at the character level rather than the word level and trained on a corpus with much longer thematic continuity (plays which are hundreds of thousands of characters long vs. songs which with an average token length in the low hundreds). The training success, then, is likely due to a simple memorization of the corpus, exacerbated by the use of the same batches in the same order in each training epoch. For this problem, non-stateful random batch selection is likely a stronger strategy. Still, the use of statefulness in the simple model does not seem like enough to explain the scale of the difference between train and test performance or the "lost model" with much higher generation cogency.

I do not consider this project to be completed at the current and plan to continue work on it in the upcoming weeks. I plan to start by making my evaluation architecture more robust so that performance can be more accurately monitored over time. This is important given the long training time needed to reach a plateau, and the current divergence in training and testing performance. A more robust architecture would allow for better monitoring of testing performance and more robust hyperparameter tuning. Additionally, I want to explore changes to the architecture, including a non-stateful LSTM layer (this would probably make more sense with random batch selection), different activation functions and decay schedules, and the addition of dropout.

Once this iteration of the model is optimized, the project can be expanded by adding more features to the current formulation (e.g. parameterizing by sentiment), or expanding to different and more advanced formulations like character-level generation (though this reduce the opportunity for parameterization) and transfer learning with successful transformer methods. Overall, I think there is potential to vastly improve the disappointing current results.

Appendix:

"Lost" topic-informed lyrics using simple model ("religion" topic):

"well i just cant believe that world no heaven <l> but how that shine is how to shine <l> how youd have been one on the other day <l> and with all my heart i could be could <l> oh i never done <e> singing mary wings shine at the unk <l> to the war of the close on a unk unk <s> to the unk the bar and the shell <l> and the unk and the fight is empty the unk <l> and when the day all done <l> the good lord gave my mind to you <l> and i dont have to wonder if love is made of unk <l> i can just hear us from you see the world <l> around the world to me hide <l> and around the world in my unk <l> and the tears in the eyes of the man <l> id be on my knees and i see the light <l> for the seen in the unk and the prayer in the faith <l> and the mary in the unk is where the unk all about <l> but the unk come to me <l> and the angels in this one down the last changed <s> <l> one day that i can hear happiness <l> the only one day by the night is gone <l> and the night is oh so the unk gonna be <l> so in my along life i do <l> for the unk in the corner of the night <l> and the unk and the unk <l> in the tree of the church <l> with the unk unk and the unk of unk <l> in the takes the people on the river <l> my unk fight and the unk is a wheel <l> my unk like my unk and an angel and my heart is in my prayer <l> though my life is my god <s> chorus <e>"

Lyrics trained using "advanced" model:

Standard Generation:

"you couple and moon <l> <l> <l> <l> <l> <l> a unk unk band <l> <l> <l> let <l> ah unk <l> <l> unk ah unk unk unk unk west <l> unk god of unk unk unk unk unk eat unk unk west unk unk at at ready to so a the kiss be can the got waitin to mom yes what side of tree pretty what early the hope the really the really like have and wife wife the took people in heart as the the was was was was said to when you you i were all all all all all all all all all all all southern and sad <l> <l> <l> <s> <s> and loves <s> went all an <s> knows my all you all all me until that the breath see a blame just a have a hell her a find very just that taste i even its unk only any me its me bed it me dreams she under i feel for as as bill i see in son again at to makin lord you shoulder were is is i know at under but sitting to i play beautiful men back are how yourself ooh our the loved the don't my speak like like go got made for unk same til have and lot to and million face or inside round the have party the pull and different as as be of i didnt with see hear same or have than as my happened until or be well have skin unk unk one of i phone to too times before so a a me texas and picture and guys to the had only who the was was had to but joe to so burning to yes i did even i are the <s> and bitch <s> <s> i i i was was will holdin but <s> <s> <s> <s> <s> <s> <s> dont you he its out ill your believe <l> <l> meet on me my you later me red or roses be grow grow grow on anyone on and baby ask work unk ball of i gave in beside unk story was unk unk unk future unk dog i took in dad the the thank the listen oh and second one of same came came look us back that the hanging when not a world unk school <l> us hearts my my never my never my doesnt <s> without be these like like are unk tall who is or ever beside wait not a you be the unk world proud of i see trying of road school these change treat doin down from and unk place i see lost inside here me black with even don't were to uh gonna sleep about the ever that that that you must i trying unk near air that the win it my but hurt but all the never glad much ive is too too a make in way you looks you you on you i might in way heart we unk king close him unk unk unk low set i know and unk earth of unk act you the must too unk same kind <l> <l> <l> <l> when with see doin in money asked in fear in mine takes us check kind i into i know all a wheel before talking you put i have far ah any my should my doesnt i too floor never wear fit in friend someone always my to don't not to goes is my comes had good me your under to love love love love love love love love love love love love love love love love love love love love love love love love love love unk

unk unk unk unk unk first white your an to your not you that the dog for my unk same <l> <l> always em to to <l> my always a your a the holdin your a blame when just its a one of i should my <s> <s> <s> <s> <s> <s> <s> <s> <s> <s> <s> <s> <s> love little it it the <l> now now come mine im lips to here ill my takes me the speak unk way night been here sand but with me me the get everything cried cold <l> johnny <l> <l> <l> <l> a river verse unk sea 2 unk war white unk one mine to everyone on in name <s> gave the will in father <s> <s> <s> mr <s> <s> <s> then baby does you roses first i first sweet me unk unk met my happened me the unk father of you you we it true <l> a the trust em to longer i crazy sweet your your in wife love kid <s> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> children aint unk fear of always me a a a a ride really plan changed to unk asked word is right hes room the must on without as right again job tv and blue long day week more unk time on on is i down is the don't the said in peace in well kiss with believe because i may let bring the leave and couple <s> put you let stop the sitting to unk working light <s> and guy high hole stop and win and only aint there me the walked it luck the know to my went like gettin pay for crazy crazy the blow my and crowd tried <s> now be ready when to but rich nothin but a bring young move i move i have make it a be the wanna them to summer honey empty but burning a the forget up many up they to i turn mom are my unk only said to unk same playing over black i seems <s> shame the was was was was said <s> <l> <l> <l> <l> <l> <l> and unk unk unk unk unk unk unk unk unk unk unk unk unk unk unk luck his"

Topic-Informed Generation:

"you couple and moon and unk unk unk <l> <l> <l> <l> a the get unk dog the get train you sure you tree a city <l> and work wife a win <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> unk unk unk unk unk unk unk unk unk earth they to to <l> unk unk unk unk unk unk unk same <l> of <l> the rollin unk unk house and fucking in lady a that his many that the a unk unk unk house <l> and little a get joe the drop <l> <l> <l> the unk plan <l> it it unk unk drunk i <s> <s> <s> <s> <s> <s> <l> the the was was aint like unk cool just a open in jeans just no the sleep you i <s> <l> him on me a me my big you i give fuck see that like pull like talking like like like like like are like are and little i <l> <l> i be i see in unk unk unk unk high and unk called and little heres you you you you i unk hanging and hell of i make up unk clothes reason <s> and little if very that the dont that the and two and head word of was said two and look you i give get i see with that that <l> that that makin <l> <l> to and unk joe unk tree your unk unk kinda cash <l> school a a the knees and unk dollar house of ol just a fire <s> <s> <s> <s> if if mans <l> and dog asked in dog <s> with know a dog <s> some and hand about me the hill my up nice the want it my be the got in unk unk unk taste i drink and long long long i a break i but the was was was can down night had like eat the was was was was should to mom like are the done <l> really was was was was was was said to truck when as and party and six and unk unk unk unk unk unk unk unk mom unk bar <l> truck <l> <l> <l> <l> <l> <s> <l> <s> <s> <s> <s> the <s> <s> <s> my the was was was said on me your with me me the be the grow a you have dirty i are in unk boy tree my so all you you i i unk unk unk unk and and and good i i said and funny but funny to now i talking this you i say a change little have may in little <l> <l> <l> <l> and drinking unk unk unk unk unk low a a a unk fun you a the unk unk unk unk taste of i say the unk unk unk unk unk unk unk unk mom i and crazy i i and little i dont too too too holdin holdin spend us unk unk unk unk story or me unk unk <s> <s> and unk unk story or have than in than in than in pay stop <s> and little <s> and little <s> and little <s> and little that the get unk unk tree tree our <s> and little little <l> <l> <l> <l> i <l> <l> <l> i <l> up i <s> <s> <s> to and little in head the thought the would be on in little little my my was was aint my <l> <l> <s> some <l> it it it it it it unk unk unk unk unk unk unk unk unk unk unk unk first to hello your better to your sorry it my fucking and million <s> and little for for hit the dont the was was was was was said <s> <s> i <s> the dont it a drive in head <l> it it it it it the and little red <l> <s> i didnt and bed <s> and little <s> and half <s> <s> and blue band was was was was had like was learned with a eat a my <l> to and

unk tree unk hole unk unk unk unk unk then then what long and unk bit the cause be the like like get unk hot the take holdin second your blue tv <l> <l> <l> <l> <l> <l> <s> <l> unk unk glass <l> in family unk unk unk unk unk unk unk unk unk unk unk unk unk unk unk unk unk unk unk unk unk unk unk unk unk unk unk unk cowboy <l> that that the and unk face and and and and white the white <l> and unk unk unk unk unk unk lady and what for for for for about <l> <l> <l> <s> <s> <s> <s> <s> <s> <s> <s> <s> <s> <s> it unk reason unk <s> unk little then like gets the turn like work in wife to unk little i unk unk unk mans <l> the ever to the was was was was said to can to as as <l> i see there go water my on my might my got unk unk unk lady in the the was was was talking to good you the met you i check i the tell to all the was was was was unk back <s> and little <l> <l> <l> <l> the be up church <l> <l> <l> i and unk unk high one of long night got on you unk devil got unk unk name i unk unk boy i got to but you you i are lord how unk luck and little unk unk unk unk unk unk cause of unk unk cause eat i gas me and unk unk unk help and little the and n i was unk unk unk unk unk unk unk unk unk boys unk unk black i <l> <l> <l> <l> <l> <l> <l> <l> and little old on and little and little the am”

Expanded Vocabulary Generation using Embedding Space:

“*you porch those <l> and tungguki unk unk unk aboutn unk kinda unk unk unk unk unk lord <l> <l> and luck <l> <l> and meka baebae unk unk tyt unk west that knows the everything to and little your of hot see women is i and little <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> up <l> why unk unk verything unk unk unk unk unk unk 記号 むにゃむにゃ burdur unk ヘルマン unk unk unk unk maduroy pumper unk unk unk unk unk ウィルソン unk talknn unk unk mom figure unk kurumunu unk look <l> n and different what did what what <l> <l> <l> <l> <l> <l> <l> <l> <l> they mngeluh unk abou unk jibawi unk itnnif commanding unk unk unk unk unk unk country unk like what trust mean <l> me the unk unk run stuck on in luck me the ibo unk unk first unk unk lhr unk nort unk unk lady that that does my my rev unk unk 玉森 unk picture <l> i thought in eye in yeşilpınar unk unk and luck <l> kartik unk couple ery afra unk aine unk hell unk unk int unk unk unk waitnni unk baeutiful fast women <l> that it that some boys <l> they from shot and picture and scary boys in tv in heart some in tickets are very on <l> can those <l> <l> me all me upstairs pay where some it women unk jack had story may the had unk unk unk blo waitnni unk unk unk alek unk unk unk eat the really or friend you my as <l> <l> but in unk nails in check in unk look and turn like get kick save in eyes saw the in last turtle might <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> □ˆ unk unk unk unk unk unk terr unk gej unk ery unk unk memon unk unk leach unk unk unk unk adt kooley riv neek chy unk garden syh unk many thought to <l> まだ起きてるの cash radio and what the and unk unk land story to to ats good you think your heres bound come a get my the was was was was ever to all all you i were sure i all all its dtw those name in brown <l> <s> <l> <s> <s> for the was was told to but where way a talk the name all all this this at been then at for a feel <l> jones expect go go came of i may <l> would be go to back <l> <l> <l> you unk ibe unk nipsey unk unk fek unk cowboy that lot the and dog to and unk house of fucking unk cause cause unk dog i イヤフォン unk レーン lf what i but the and shit you i are the needs the still taking hello well the of am of was was said to so a me the told and hair <l> it it the do kinda the dont the saw and little all all the as the forget the its the call and little little i come i see in all a a down i a with take right and man just hundred band and darlin get mean unk red of i come a be of i not <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> <l> yea <l> unk elly unk unk engkaupun unk unk stephens weebie unk tollway ery unk followinq urslf unk chirinos nyman unk unk unk unk kayne unk unk engkaupun unk unk unk unk unk unk unk making <l> that a of unk advani unk unk wr unk adnan cowboy an i to and upstairs in land you lucky a play em harshvardhan lot with was saw to think not your with unk witt boy either some to <l> to kevin on joe on unk unk brsma unk ery waseem unk world the stopped to the made to like have kelly the shake a the unk* *with and boy dog the old of ass get i smile of cat up you my agree started im for i working i dog i pick blue all a so on in young in unk ery unk unk vajpayee endustri unk dup unk unk unk hello unk shit we they unk unk african about little i and hanging if the vontae northbound youncare おまえだったのか unk unk unk unk unk and tree from with this and ten and young you you you you you you you mean i make a dog uwa whole of i dont in lips in life then i thinks back <l> <l> up <l> <l> <l> <l> <l> <l> brdg unk decarie kind too unk shafika unk tahir pop vajpayee unk tj the drink to jerry in the the the might to <l> <l> around <l> blo unk unk unk unk unk unk aspherical fek caz unk cowboy they to but your brsyukurlah phe unk unk suh tree clo unk tree unk meant how what <l> the the <l> them on on in unk tree your in in drink in <s> <s> for some some <l> <l> from walked you of could make it blame <s> <s> ahh first the dnt instead a you i fucking <l> had a a me the on a you little i see all all friends your capt unk chulilla long guys to forget i always guy me your different the win two the throw the unk unk money and tungguki boy タクト part unk unk kampusu 武田 rain i unk unk unk first i in unk track eye but ilce unk body unk unk unk kash bit you i and little how i with some i and i and unk one i i and i but stuff too way to so so a me it me this i see with be"*