Machine Learning Engineer Nanodegree

Capstone Project by Anshuman Sahoo

Learning the lyric writing style of Bob Dylan

---

1. Definition

1.1. Project Overview

"Take care of all your memories. For you cannot relive them."
- Bob Dylan

Or can you? Natural Language Understanding has developed into a diverse and important component of the AI mission over the past few decades. And understanding sentiment, tone and writing style of textual data has become an exciting application of this body of work. This includes models that filter out terrorist chatter on Twitter, classify positive and negative movie reviews and auto correct emails.

Lyrics contain patterns of styles and sentiment that persist across the work of artists and sometimes across artists within a genre. It is a seemingly fantastical idea that someday we may be able to take elements of artistic talent from our favorite artists and create an agent that produces lyrics that retain those elements. This is the challenge chosen as the problem for this capstone project - to try to learn the styles of artists and generate lyrics based on those styles.

Recurrent neural networks are the most popular method to learn sequences such as sentences. They work by allowing the computation to be influenced by the past computations in addition to the current. They have been used in various situations as described below.

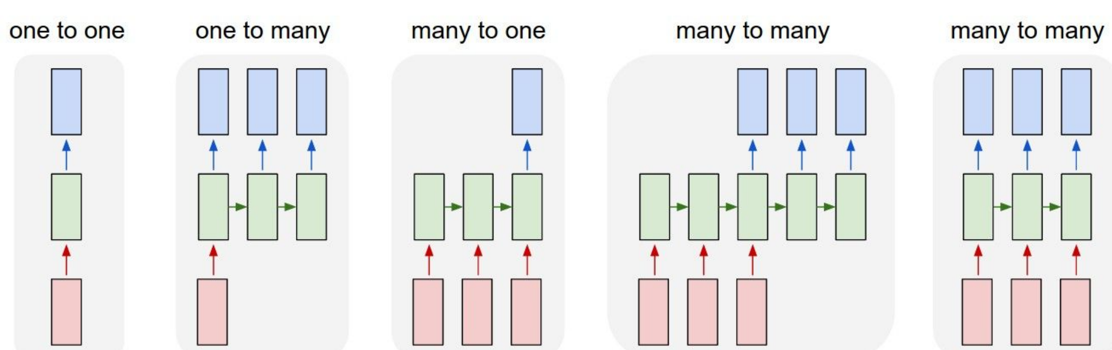


Fig1. Each rectangle is a vector and arrows represent functions (e.g. matrix multiply). Input vectors are in red, output vectors are in blue and green vectors hold the RNN's state (more on this soon). From left to right: (1) Vanilla mode of processing without RNN, from fixed-sized input to fixed-sized output (e.g. image classification). (2) Sequence output (e.g. image captioning takes an image and outputs a sentence of words). (3) Sequence input (e.g. sentiment analysis where a given sentence is classified as expressing positive or negative sentiment). (4) Sequence input and sequence output (e.g. Machine Translation: an RNN reads a sentence in English and then outputs a sentence in French). (5) Synced sequence input and output (e.g. video classification where we wish to label each frame of the video). Notice that in every case are no pre-specified constraints on the lengths sequences because the recurrent transformation (green) is fixed and can be applied as many times as we like. (source : karpathy.github.io/2015/05/21/rnn-effectiveness/)

LSTM or Long Short Term Memory networks are special cases of RNN which have the ability to remember values between arbitrary intervals and hence are suitable to learn patterns in arbitrary sequences.

1.2. Problem Statement

Hence, in this project the goal is to use a Lyrics dataset from Kaggle (https://www.kaggle.com/mousehead/songlyrics) to generate a model that produces lyrics that sound like that of legendary singer songwriter and folk artist Bob Dylan. I decided to use a model used for character level generation using LSTM and fully connected layers. The goal is to generalize character prediction without seeming to memorize the text to generate unique sounding lyrics.

1.3. Metrics

I decided to use the number of erroneous words generated/total number of words as a measure of how the model is doing. I used dictionary.com as a resource to verify if a word exists.

1.4. Long Short Term Memory Networks

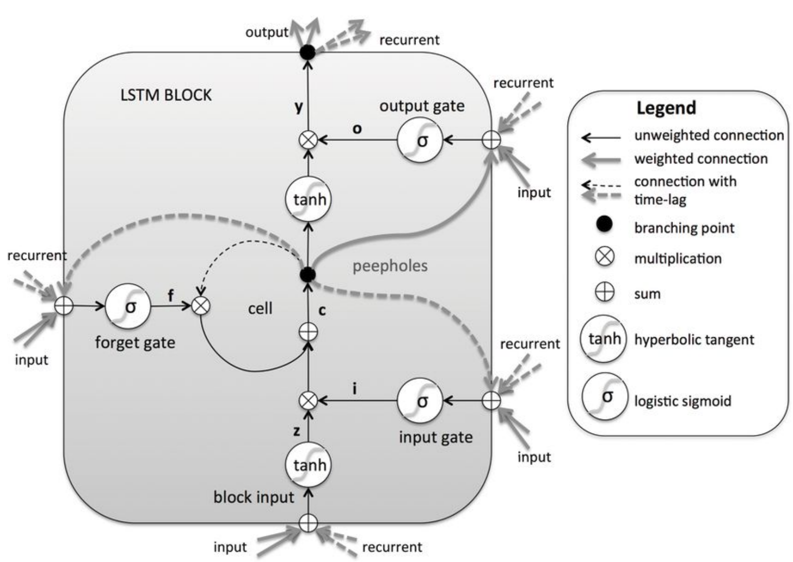


Fig2. An LSTM block (source: https://www.researchgate.net/figure/292303547_fig8_Fig-2-Long-Short-Term-Memory-recurrent-neural-network-architectureA-single-memory)

RNN's should theoretically be able to learn arbitrary sequences given enough data. However, in practice they face a lot of issues as the interval in the sequences grows. Hence, LSTMs were developed to be able to learn these long term dependencies more easily. They have been designed so that retaining memory for a long time is their default behavior. Each LSTM block in fact consists of four neural network layers that interact in a very particular way in order to allow some data to pass through without changing the state of the block while altering it in some cases.

LSTMs are one of the more popular architectures used in natural language processing today. They have variants within this architecture such as one used in the project called Bidirectional LSTM , which allows for the information to propagate in two directions across a network.

2. Analysis

2.1. Dataset Exploration

The dataset (https://www.kaggle.com/mousehead/songlyrics) consists of a 72MB csv file with columns named 'artist', 'song', 'link' and 'text'. This is one of the smaller datasets available in this category and I chose it due to ease of computation with limited hardware.

A snapshot of the data is shown below:

artist	song	link	text
ABBA	Ahe's My Kind Of Girl	/a/abba/ahes+my+kind+of+girl_20598417.html	She's just my kind of girl, she makes me feel fine Who could ever believe that she could be mine? She's just my kind of girl, without her I'm blue And if she ever leaves me what could I do, what cou
ABBA	Andante, Andante	/a/abba/andante+andante_20002708.html	Tread lightly on my ground Andante, Andante Oh please don't let me down Andante, Andante Oh please don't let me down
ABBA	As Good As New	/a/abba/as+good+as+new_20003033.html	As good as new, thank God it's true Darling, we were always meant to stay together Yes the love I have for you feels as good as new Darling, we were always meant to stay together
ABBA	Bang	/a/abba/bang_20598415.html	Bang, a boom-a-boomerang Dum-be-dum-dum be-dum-be-dum-dum Oh bang, a boom-a-boomerang is love A boom-a-boomerang is love
ABBA	Bang-A-Boomerang	/a/abba/bang+a+boomerang_20002668.html	Bang, a boom-a-boomerang Dumb-be-dumb-dumb be-dumb-be-dumb-dumb Oh bang, a boom-a-boomerang is love A boom-a-boomerang is love
ABBA	Burning My Bridges	/a/abba/burning+my+bridges_20003011.html	Moving at last Girl I'm leaving and I'm burying the past Gonna have peace now You can be free No one here will make a sucker out of me
ABBA	Cassandra	/a/abba/cassandra_20002811.html	I only saw it as dreams you would weave Until the final hour I'm sorry Cassandra I'm sorry Cassandra

Fig 3. Screenshot of songlyrics.csv

Statistic	**Value**
Number of artists	643
Number of Bob Dylan songs in dataset	188
Shortest Bob Dylan song	I'll Be Home For Christmas (79 words)
Longest Bob Dylan song	Bob Dylan's 115Th Dream (793 words)
Average word length of Bob Dylan songs	254.675 words
Average word length of all songs in dataset	219 words

Most commonly used words by Bob Dylan	[('the', 2033), ('I', 1516), ('to', 1017), ('you', 886), ('a', 886), ('in', 661), ('and', 649), ('my', 636), ('of', 635), ('And', 499), ('me', 436), ('on', 390), ('your ', 382), ('that', 333), ('for', 319), ("I'm", 310), ('it', 305), ('be', 304), ('is', 288), ('was', 260), ('all', 253), ('with', 220), ('no', 208), ('But', 206), ('got', 204), ('down', 202), ('they', 200), ('You', 192), ("don't", 187), ('up', 182), ('Well,', 179), ('just', 178), ('The', 175), ('like', 174), ('can', 172), ('from', 163), ('her', 162), ('he', 160), ('know', 160), ('are', 153), ('have', 151), ("it's", 142), ('what', 138), ('his', 135), ('so', 129), ('one', 127), ('do', 127), ('When', 125), ('out', 123), ('will', 120)]

Table 1. Dataset statistics

A lot of Bob Dylan's songs are in the first person, and this is evident from the most common words used by him. In addition, he writes songs slightly longer than the average song length.

2.2 Environment

The code for this project was written in the Keras using the Tensorflow backend. I have a 2GB GeForce GTX 1050 GPU which helped to speed up the training process a little. The packages were installed using Anaconda. Other major libraries I used were Pandas and Numpy for pre processing the csv file and preparing training, validation and testing sets.

2.3 Data preprocessing

The csv file was read as a pandas DataFrame using the read_csv function. The list of artists were filtered to pick out the Bob Dylan songs and the corresponding text is appended to a cumulative string containing all the songs.

Total vocabulary = 74

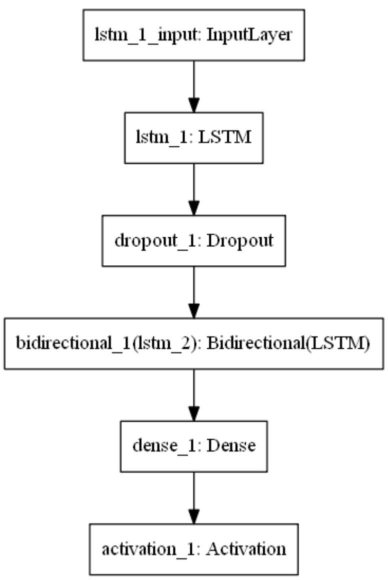
Corpus length = 260264

The text was then cut into semi-redundant sequences of 120 characters. We need to convert the characters into sequences based on character indices in order to feed it into a neural network. That is, we will have sequences consisting of character indices from 0 to 73.

The next step is to convert the sequence of character indices into one hot encoded vectors. This helps the model to be able to predict the probability of each of the 74 characters in the vocabulary.

3. Methodology

3.1 Implementation



Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 120, 128)	103936
dropout_1 (Dropout)	(None, 120, 128)	0
bidirectional_1 (Bidirection	(None, 128)	98816
dense_1 (Dense)	(None, 74)	9546
activation_1 (Activation)	(None, 74)	0

Fig 5. Model Architecture

The model consists of an LSTM layer with 128 memory units followed by a dropout layer that randomly ignores 25% of the neurons. This is followed by a Bidirectional LSTM layer. The reason to use multiple LSTM layers is to capture structural elements of the text and style that might be overwhelming to capture using a single layer.

Bidirectional LSTM layers are better than normal LSTMs in capturing context because the flow of information in a unidirectional LSTM is such that it retains only information about the past. However, bidirectional LSTMs are able to provide predictive powers to the layers by providing future intuition as well. I hoped that having a bidirectional layer would allow me to capture the structure of Bob Dylan's writings.

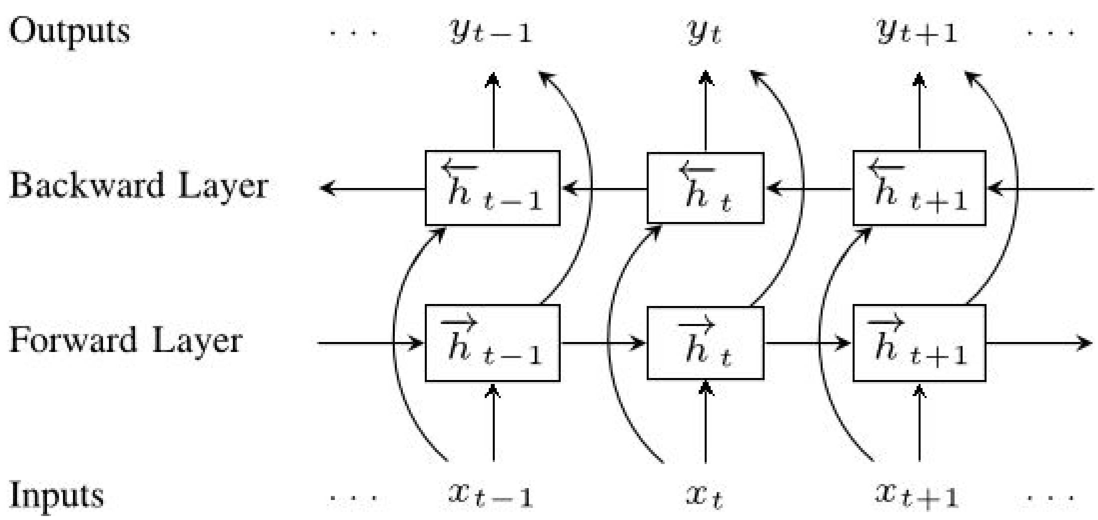


Fig 6. Bidirectional LSTM

Finally, we have a fully connected layer consisting of size 74, equal to the size of the vocabulary, with each softmax activation outputting the probability of the next character being of that index. The model is compiled using RMSProp optimizer and a loss function called 'categorical_crossentropy'.

3.2 Training

The training process ran using the iterations method where the model.fit function is called iteratively and the number of epochs is fixed to 1. This allows us to test how well the model is performing qualitatively at various points of the training process. I ran this iterative method for 100 iterations and saved the outputs in a txt file at each iteration for evaluation. The learning rate is 0.001.

4. Results

In order to generate text with varying amount of diversity, I found a temperature based sampling method that samples an index based on a probability array output. I generated text at every iteration.

Iteration ↓	Temperature 0.2	Temperature 0.5	Temperature 1.0	Temperature 1.2
1	oo le we we the ou o o he he ton . . . I I o I oo I o le the ou	me tou I Ae aore o' milo ou the lat or to he tore mo oie I ue I the . . . mee We foug ther o o the T ton le ole t	i, yiom te, oe dhe, theriaft c we boek hdg' bAteit sheioe, to I fere hhvt oD'm p'm, on gy'n s haua yo rlngawd . . . I sint ee no oteeuhe ao, wanr sidy out mo rodh nhm hiu' tbYis, bokl ius mer shee Tpr lered, Whe tee, tee thlecr,y aut uune vng aou ce yeta, G	vxcn ph.' Byo mrrnr'e soeJ n not on ho'be ar't te me 'aw''n clwk moud'ime Aio oo hh ul Sseeh.y 'ne yDml'r w'n so nl ,illiwe'lice miotee oTstteaon'tY yodee cYet to on sh les eot mory dlabn, mcalind pof Nele wott h ceurd shirmy cettlg fle oos las, Df,rllny Dorot hhe oo' nos fhs. elm,oso fhes ba atnr yeye oyu mon, mn fenk mf sim d f ie tha oilg ,or,d webhe,pt ch hetoein
25	oune. And the said and I said When you're be and it sead And you ready of that what you be the said. And I was a been and the been to me to see I was a came I said the carker and a muse, The stristle what I was a caus I was a can the stristan I stad the said. The said a star and a champert I was a can the said I was a clauling all the stristles The stristle	est And the saids in the cover When you some for you. I don't gat in't of my forly. I moane, it's wint the stang and he said. I was and I have to crave I'm seel a looker me If a gonead the stimes They're standy in the carsing back the said. The stripples and the same of the couring the head. Fain the load, I can't stand the stristles. Oh, I can't gave the	ace You in the ssriver on his He stancy When than peath of a chasked you made out We'le jeen evan cowne in a chimcest And I tever and'd hask and is me '" fare manes, I'm saadion, iivinglentice the spasight. Floow up the sead, Catte past me, babe, I care mine me boy Thing, been the llook that starks one, I was begit They keep, it bik shay Lord's so hurnen't gercas It's	e. Shood If's in neadin' is cloised. I won't eled say, it's handy, deelry mach a dagiidin' black. Through of's I jabyed. "Ach, bic, you. Sum, mad. Well, you dune're thilp, go, my blake, Don't king broken, you him I wast unky. He'veurd yoa man here's boung stles, my gite ligh my eazite? I tere.Yea homethere walllesboog is they chor crysele the catloren, She ost walki

50	ame you may To chan mine in the fare And it took your bod Yor up to be the whole In the way an true Liking to your lover you Well, I want to be your heart on the save you I'm up a bink you want to me any monty, And you're from the bring wing to blow. It a cut a forly on my heart And they done monked to aty to be too They's a mund to be the change can by beep I can't be **Score - 4/89 = 0.0449**	ame my feed. They say you that some find Hightay She sead broken with the door wands He was a strit to see all spipp to blowing Stonesin my caull show me angens She say to see you and here on my hard Chrost I fell it's strought I'm love When we say but I've gone my mind for you They ase you love my heart and they would selled strill And I malk around the sumpans of the song The **Score - 14/81 = 0.1728**	ome mone." They have you clyak and have the eatcick Now master apone in my man. As the sace has arrier "I've were from trikus I coll ase was done byen I'm noinat in the peardy of you got to do All he'll see on through, I been on Well, I one maby kindin' of treed She dead you get out for a done, Take a stall you suill be a caus Or for you pluce It's nit me for lot my he's **Score - 8/83 = 0.1558**	ommy lostecp The's well from him Har down Ain't like you dread, it I kneam, They was dead befled us Coplay oter blacked grin', got long, If' this just it nobedy, in hame is But them alwhy cry, Or may sackes, be men inside dowr, With from that bound dyguver, new that just cliyes?I pisbed. ome wus you comes away you cowins They're mollemon, they'ndyed anoweround feelo. The locu **Score - 6/70 = 0.0964**
75	And you had wish you dread from the cried. She was remore fur the sky, We'll me. But you read and more. I was a-cownin' for a word, They're gonna good friend. If you say to hear my money, And you want to the would new han drowd If me, I was allon you don't see. I was wall An the ard the see. When I year! It torned that you can much to blut I was sky and when y **Score - 6/81 = 0.0741**	And you got my sky bress be you. I never been by next for the city, Or said, I just gonna hear my back your cold. I been me stail, baby, I got go black you may yem, fliend. She said you pay me window, Oh my nave List a contire sherowing She's a beed and strenkin' she wing as strees. And I don't curt the saids I stom the reolle be. And a stroughtiint of mine. And I t **Score - 5/80 = 0.0625**	And have gocnard it pass be he ausade To stan. it's danning, Snever was lett my dayin' For where man the sumpare, now what a down Cowlly had liver, And the and out my cris. She love that poecout with he cove town. every tire stimed and put a still don't had bound, All dod Ditime, I seel, And we was ally to werd And you're blowing I was friend to rably ben. May I've good, **Score - 7/76 = 0.0921**	And you uren the love of Your day Their you before the bend Dywurd wike no more Oh to by sring mide" Or manace like the pary, run he gun the fuent Ig's from Choste strongs, the lass alway fas And a sempsrych fraid, "He couod mad so putte And I'm lovar that me forsuull heart flowswand Ain't back it his wolderyssleet from the way. Jughed that I'm glestit's bleayy. merure up no **Score - 15/73 = 0.2055**

100	e don't dinto to to to beet I tell I'm something in the one and hear and bell that I can't get kind When I hee them one mone mes and I'm fling and beckly behind. Hear you please chay know it all my food. Well, I want to be your houst see maby Like all your commars and and ster Say you're go the said. They, hatk your fordard Jack. You know I'm strim the look all south my back **Score - 4/81 = 0.0491**	e don't been to there It baby that I can't see? A mind many ceen a broken down, With my spite, they're but he was that shee whistleded for the deer. Not is all the broke that hear my hands If the andwile on a dick and happen stands, And who babes of the clack that, I'm sair from him and the mounstantion, That to all hat can she dadse I would srund, is his fase anywain. **Score - 7/77 = 0.0909**	e get and home the best, The wallow up so the saiding chatch They puts babes you'll lake it's eabled. Jurn'd stay for blood and gland long, though the cuts. Looking, wo say, "Feef your lover nay, I want't go the good play, you have man she lives and a criel And have doned, That's on the erglen, babe, I'd dong had gone, dreass for ever many Soust they know what whith bound. **Score - 8/73 = 0.1096**	e die in the sun, no year, Ife like you went shake whar the morn. Maybin' em, in her with on the skye ofon touch ham, Leadin' of jard. I had reckad, you can dall lot him But if you will to talll Baby, you'd sight flow you and it dog come on your lon you You'll ba got a dill walk up a yever you Maum full like you. It is allowh, there's can for you. Gen you'k word cramming **Score - 4/81 = 0.0494**

Table 2. Results

From the table above, we notice that the text generation stabilizes around the 50th iteration and we get an average error of 5 erroneous words per 80 generated. This does not mean that the model is able to learn grammatical rules or style, but is able to correctly form words.

The metric chosen is naive in comparison to context based metrics used in other papers. However, due to time constraints on this project, those explorations are set aside for future projects in this area.

4. Conclusion

The results show a progression in coherence and use of some Dylanesque words in a familiar style. However, since the character level RNN does not learn meaning but rather mapped the character relationships in a higher dimension, this project demonstrated to me the limitations of sequence based models in learning more abstract things like writing styles. However, it provided a good baseline for how LSTMs learn sequential information.

Next steps in this project would involve exploring and experimenting with Word level CNNs and Word2Vector representations of lyrics as well as hierarchical models that would retain the structure of the data and would in a truer sense capture the style of an artist.

5. References

1. karpathy.github.io/2015/05/21/rnn-effectiveness/
2. https://www.researchgate.net/figure/292303547_fig8_Fig-2-Long-Short-Term-Memory-recurrent-neural-network-architectureA-single-memory
3. https://www.kaggle.com/mousehead/songlyrics