Generating Political Monologue and Dialogue using NLG

Mohammed Sahal Ahmad, Sushma Suresh Kalkunte

November 8, 2021

1 Introduction

1.1 Summary

Some of the most important speeches are the ones that national leaders make for political reasons. They are extremely powerful as they directly impact the lives of not only the citizens of the country but also international relations with other nations and their masses as well. Throughout history, we have noticed political speeches being used to motivate the population, begin revolutionary movements, even start and end wars.

The reason why political speeches are so powerful is twofold. Firstly, the script itself is designed and written in such a way that it stirs the emotions of the audience and leaves them thinking or wanting to act. Secondly, the political leader is a great orator. He/She/They/Them deliver the speech in such a manner that every single word they say sounds convincing and makes the audience want to believe in their cause.

1.2 Problem Statement

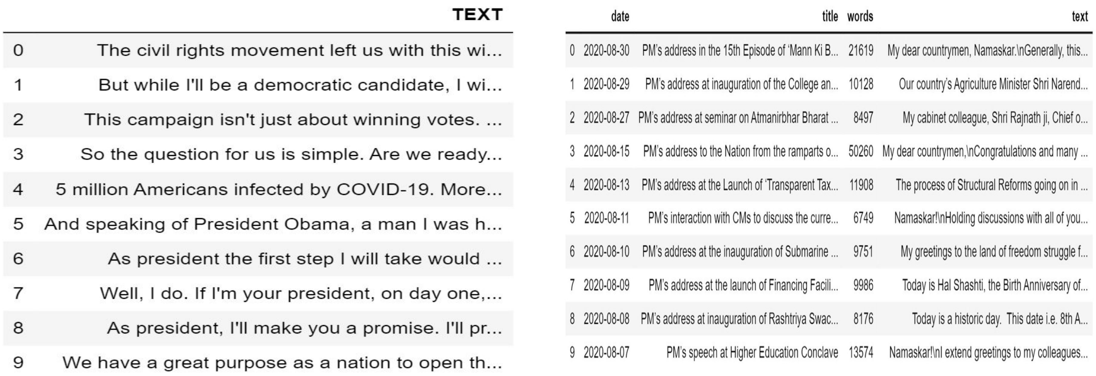
In this project, we build a system that can learn the art of writing political speeches and generate a speech that is very identical to that given by two current world leaders, namely Prime Minister Narendra Modi (current PM of India) and President Joe Biden (current President of the United States). By combining the domains of Deep Learning and Natural Language Generation, the model learns patterns and usage of words over the speeches made by a political leader in the past. It then uses this context to generate new text for the specified seed.

1.3 Data

For Prime Minister Narendra Modi, data has been partially downloaded from the Harvard Dataverse (access to the database was provided upon request) and partially scraped from the PMs official website that maintains a transcript of most important speeches. Via both of these methods, we were able to collect over 1500 text files with his speeches given at multiple occasions between 2015 and 2020.

Obtaining data for president Joe Biden was slightly challenging as he has given fewer speeches over his current tenure of less than one year. Transcripts of his speeches were scraped from an open source Audio/Video to text website. Close to 200 speeches given by Joe Biden starting from November 2020 till October 2021 were obtained in this manner.

Sample speech data of both world the world leaders are as shown below-



	TEXT
0	The civil rights movement left us with this wi...
1	But while I'll be a democratic candidate, I wi...
2	This campaign isn't just about winning votes. ...
3	So the question for us is simple. Are we ready...
4	5 million Americans infected by COVID-19. More...
5	And speaking of President Obama, a man I was h...
6	As president the first step I will take would ...
7	Well, I do. If I'm your president, on day one,...
8	As president, I'll make you a promise. I'll pr...
9	We have a great purpose as a nation to open th...

(a) President Joe Bidens Data

	date	title	words	text
0	2020-08-30	PM's address in the 15th Episode of 'Mann Ki B...	21619	My dear countrymen, Namaskar.\nGenerally, this...
1	2020-08-29	PM's address at inauguration of the College an...	10128	Our country's Agriculture Minister Shri Narend...
2	2020-08-27	PM's address at seminar on Atmanirbhar Bharat ...	8497	My cabinet colleague, Shri Rajnath ji, Chief o...
3	2020-08-15	PM's address to the Nation from the ramparts o...	50260	My dear countrymen,\nCongratulations and many ...
4	2020-08-13	PM's address at the Launch of 'Transparent Tax...	11908	The process of Structural Reforms going on in ...
5	2020-08-11	PM's interaction with CMs to discuss the curre...	6749	Namaskar!\nHolding discussions with all of you...
6	2020-08-10	PM's address at the inauguration of Submarine ...	9751	My greetings to the land of freedom struggle f...
7	2020-08-09	PM's address at the launch of Financing Facili...	9986	Today is Hal Shashti, the Birth Anniversary of...
8	2020-08-08	PM's address at inauguration of Rashtriya Swac...	8176	Today is a historic day. This date i.e. 8th A...
9	2020-08-07	PM's speech at Higher Education Conclave	13574	Namaskar!\nI extend greetings to my colleagues...

(b) Prime Minister Narendra Modis data

Figure 1: Sample of collected political speech data

1.4 Modelling

To achieve our goal of political speech generation, we have selected to implement the following popular and state of the art language generation from scratch-

1. Markov Chains
2. Character Recurrent Neural Network (Char-RNN)
3. Long Short Term Memory (LSTM)

These will be discussed further in the next section.

2 Methodology

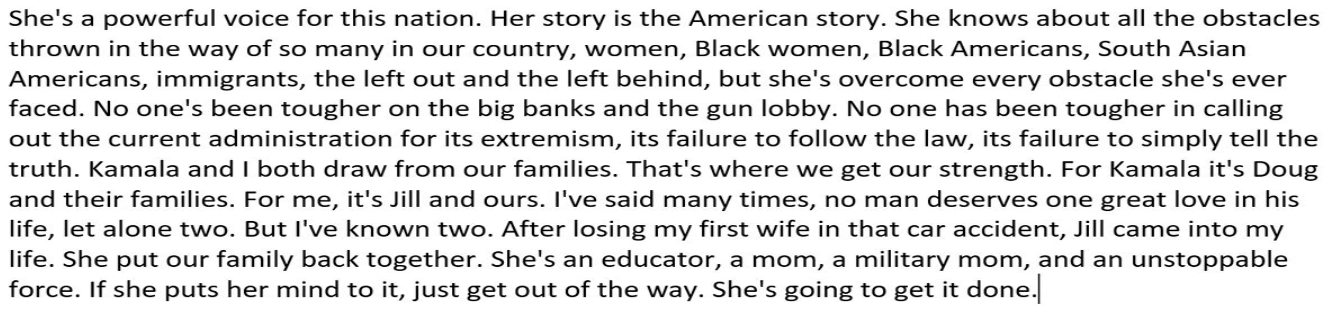
Once the data has been collected, our methodology can be broken down into the following key components and tasks.

2.1 Data Preprocessing

The below mentioned steps were followed for data cleaning and preprocessing-

1. Removing punctuation marks
2. Removing special characters and URL
3. Converting each speech to lower text
4. Correcting misspelled words
5. Removing stop words
6. Stemming and Lemmatization

The data obtained after parsing using a web crawler was not clean. It contained a lot of non-English characters, unwanted time stamps, and words that were were spelt incorrectly. These had to be removed along with special characters and punctuations.



She's a powerful voice for this nation. Her story is the American story. She knows about all the obstacles thrown in the way of so many in our country, women, Black women, Black Americans, South Asian Americans, immigrants, the left out and the left behind, but she's overcome every obstacle she's ever faced. No one's been tougher on the big banks and the gun lobby. No one has been tougher in calling out the current administration for its extremism, its failure to follow the law, its failure to simply tell the truth. Kamala and I both draw from our families. That's where we get our strength. For Kamala it's Doug and their families. For me, it's Jill and ours. I've said many times, no man deserves one great love in his life, let alone two. But I've known two. After losing my first wife in that car accident, Jill came into my life. She put our family back together. She's an educator, a mom, a military mom, and an unstoppable force. If she puts her mind to it, just get out of the way. She's going to get it done.

Figure 2: Sample Joe Biden data after cleaning and pre-processing

2.2 Modelling

As stated above, we implement the following algorithms for text generation.

2.2.1 Markov Chains

One of the classical approaches used for text generation (found in initial messaging apps in smartphones) is Markov chains, based on random probability distribution. Its quick approach is possible due the Markov property: the probabilities of future states are not dependent upon the steps that led up to the current state. This means that the process is dependent only on the previous state (previous word) to predict the next state (next word). The probability of "hopping" or transitioning from one state to another is determined by the data encountered in the corpus. The model conveniently "forgets" the preceding words/states, and this helps the model run extremely fast.

The Markov process requires creation of a transition matrix from the data corpus. This transition matrix shall house the probability of transitioning between states. The model will have to traverse the entire corpus of data and store all transitions from a particular state to the next word. The transition matrix is comprised of the word sequences being stored in rows and the next word being stored in the columns. The count of all pair wise states is calculated, converted into probability of occurrence and then stored in the transition matrix.

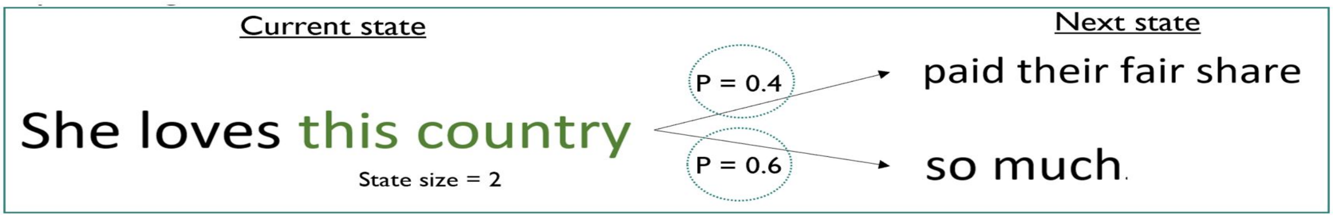


Figure 3: Understanding the working of Markov Models

Consider the schematic of Fig 3: The text "She loves this country" is treated as a current state for which the Markov model must predict the next set of words. In the current iteration, we consider the state size to be 2 and the current state is "this country". The Markov chain has to predict the next set of words for the state. Based on the transition matrix, the algorithm sees two possible states: "paid their fair share" and "so much". On account of the random distribution based on their assigned probability of occurrence from transition matrix, the Markov chain shall suggest the next state.

The Markov chain requires a "seed" or starting state to be provided. The algorithm then begins chaining from this start state, adding words from transition matrix based on their probability distribution and finally stops when it encounters an end state.

2.2.2 Char-RNN

RNN are a a class of neural networks that allow previous outputs to be used as inputs while having hidden states. They are well-suited for problem domains where the input (and/or output) is some sort of a sequence - time-series financial data, words or sentences in natural language, speech, etc.

The term “char-RNN” is short for “character recurrent neural network” and is effectively a recurrent neural network trained to predict the next character given a sequence of previous characters. Char-RNN as a classification model that includes full back propagation learning with Adagrad optimization. In this technique, we wish to output a probability distribution over character classes, i.e., a vocabulary of characters. The model is given characters one at a time and only expected to predict an output after the last character. Each character is typically encoded as a one-hot vector, where the position of the one indicates the position of the character in the vocabulary. Since this is a classification task, where the output classes are characters in the vocabulary, we use the standard categorical cross-entropy loss to train the model.

By iteratively minimizing this loss, we obtain a model that predicts characters consistent with the sequences observed during training. This is a fairly challenging task, because the model starts with no prior knowledge of the language being used. Ideally, this model would try to capture the style of the world leader.

2.2.3 LSTM

Improvement in text generation were witnessed when moving from traditional, stochastic Markov models to Recurrent Neural Network (RNNs). However, RNNs suffer from short-term memory as they lack performance if the input sequences are long. RNNs also had issues where for deep layers, it became extremely difficult to tune parameters of earlier layers (this phenomenon is known as Vanishing Gradient). Hence, to fill the gap, a modification of RNN called Long Short Term Memory (LSTMs) were introduced. With LSTM, neural networks start having long duration dependencies, and are now one of the advanced techniques used for text generation.

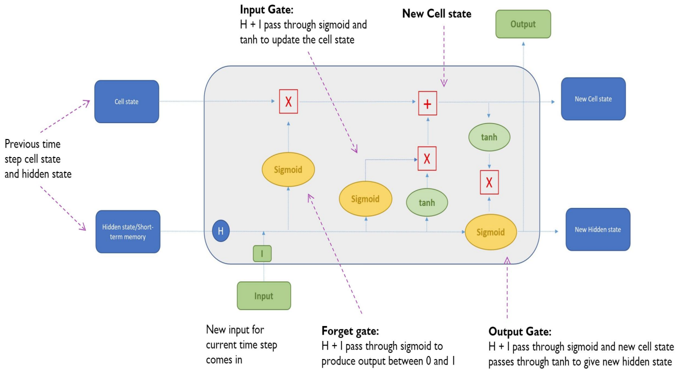

Input Gate:
H + I pass through sigmoid and tanh to update the cell state
New Cell state
Output
Cell state
X
+
New Cell state
Previous time step cell state and hidden state
tanh
Sigmoid
X
Sigmoid
tanh
X
Hidden state/Short-term memory
H
Sigmoid
New Hidden state
I
Input
New input for current time step comes in
Forget gate:
H + I pass through sigmoid to produce output between 0 and 1
Output Gate:
H + I pass through sigmoid and new cell state passes through tanh to give new hidden state

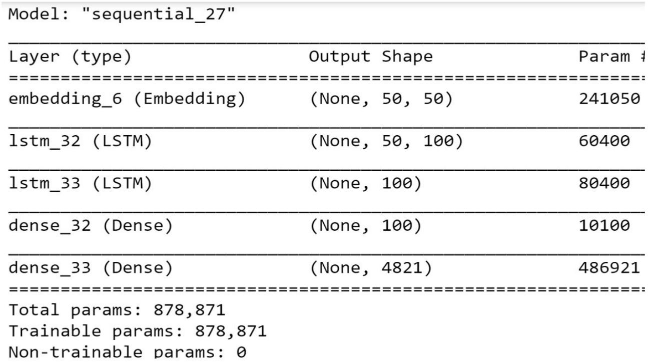
Figure 4: Understanding LSTM working

LSTMs operate by a procedure of gates and cell states which will restrict/control the flow of data/information. These gates “learn” the important information and pass only those through the gates which it wants to remember and closes what it wants to forget. In any typical gate, the new input and previous state are combined to form a single input vector. This input vector now has information about the current input and previous inputs. This vector passes through activation function and the output is the new hidden state, or memory of the network. The information passage in a cell is performed by three gates: forget gate, input gate and output gate. The forget gate is responsible for this process of “keep” and “forget”. The output from sigmoid activation function (between 0 and 1) which is close to 0 shall be the information it wants “forgotten”. Other outputs which are closer to 1 shall be “kept”. Another set of input and previous hidden state are passed through tanh activation function in order to regulate the network. The output of tanh is multiplied pointwise with output of sigmoid to keep/forget information. These combined outputs are then added to output from input gate to finally update the cell state. Post this, we arrive at the output gate. Here, the new cell state is passed into tanh activation. The previous input and hidden state are passed into sigmoid function. The outputs are combined to provide a new hidden state. The new hidden state and new cell state are passed to the next cells in next time step.

Text generation is a kind of supervised learning approach. Here, we take sequences of words/characters as input vectors (X) and the next word or character as label.

Language models require sequence of words as input data, as predicting the next word is the aim of the model. This is enabled by tokenization step. Tokenization is the process of extracting tokens from the text/corpus of training data. Once this is obtained, then entire corpus is transformed into sequence of tokens. The LSTM layer takes as input a sequence of words, where each word is represented by one-hot vector.

Once our model is trained and ready, we input a seed text and generated a speech to follow the seed word. A hyperparameter for the LSTM involved a temperature/diversity factor (value ranging from 0 to 1.5) where values closer to 0 provided more conservative results and higher values predicted surprising, even shocking results.



```
Model: "sequential_27"
_______________________________________________________________
Layer (type)                 Output Shape              Param #
===============================================================
embedding_6 (Embedding)      (None, 50, 50)            241050
_______________________________________________________________
lstm_32 (LSTM)               (None, 50, 100)           60400
_______________________________________________________________
lstm_33 (LSTM)               (None, 100)               80400
_______________________________________________________________
dense_32 (Dense)             (None, 100)               10100
_______________________________________________________________
dense_33 (Dense)             (None, 4821)              486921
===============================================================
Total params: 878,871
Trainable params: 878,871
Non-trainable params: 0
```

Figure 5: LSTM Architecture

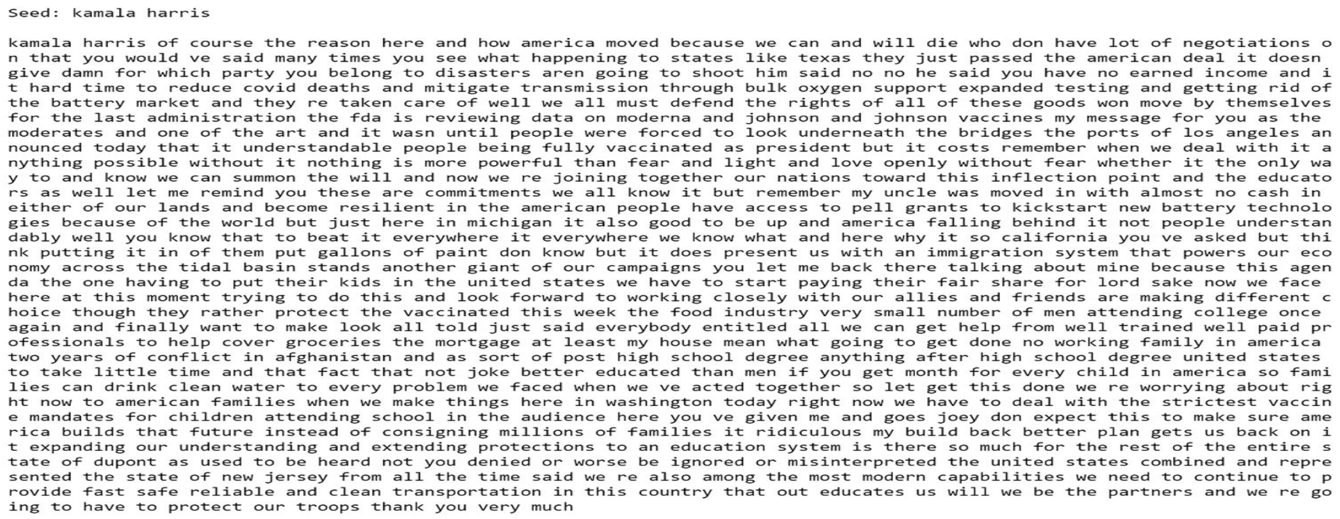
LSTM layer description:

- Input Layer – Takes the sequence of words as input
- LSTM layer – Computes the output from embedding input layer using LSTM units. The number of units in the LSTM layer can be fine tuned.
- Output Dense Layer – computes the probability of the next best word as the output

3 Results and Discussion

After implementing the techniques as described above on our dataset of political speeches, and training, we were able to generate sequences from the model by feeding an initial sequence as the seed, and then repeatedly sampling from the model.
Samples of result obtained from each of the models is as shown below.

First we trained the Markov chain model on President Joe Biden’s data and obtained the following generated text. We created three Markov chains of incremental state size (state size 2,

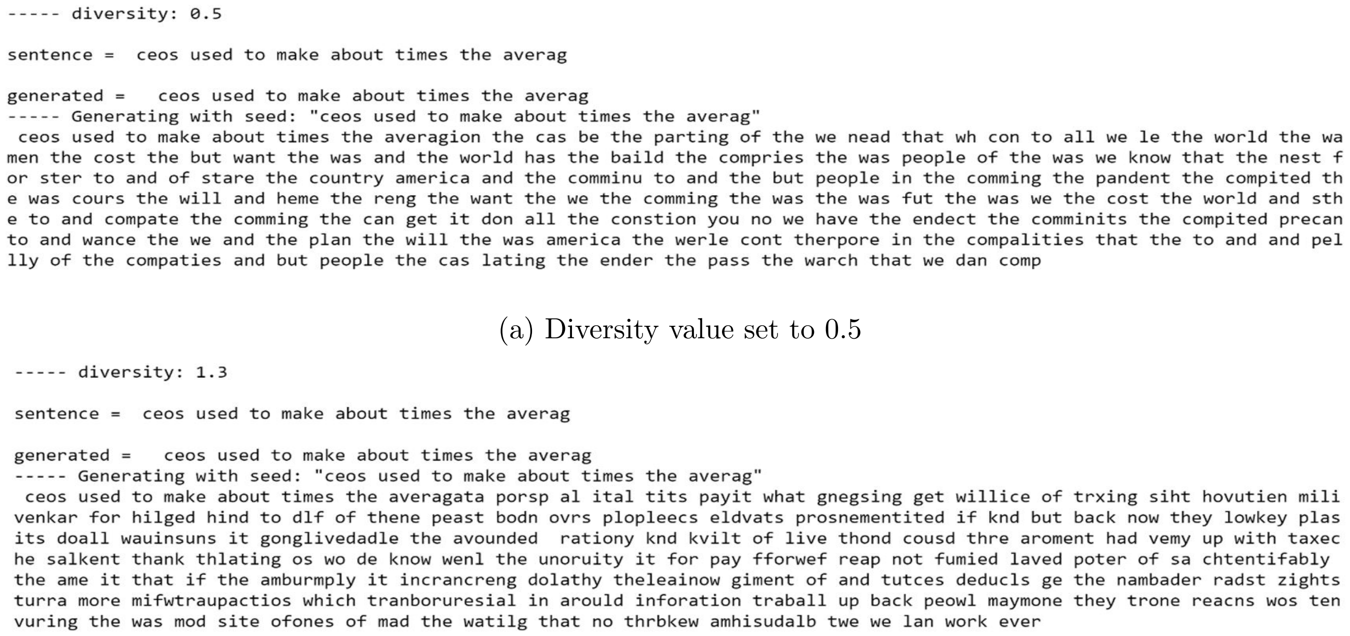


```
Seed: kamala harris

kamala harris of course the reason here and how america moved because we can and will die who don have lot of negotiations o
n that you would ve said many times you see what happening to states like texas they just passed the american deal it doesn
give damn for which party you belong to disasters aren going to shoot him said no no he said you have no earned income and i
t hard time to reduce covid deaths and mitigate transmission through bulk oxygen support expanded testing and getting rid of
the battery market and they re taken care of well we all must defend the rights of all of these goods won move by themselves
for the last administration the fda is reviewing data on moderna and johnson and johnson vaccines my message for you as the
moderates and one of the art and it wasn until people were forced to look underneath the bridges the ports of los angeles an
nounced today that it understandable people being fully vaccinated as president but it costs remember when we deal with it a
nything possible without it nothing is more powerful than fear and light and love openly without fear whether it the only wa
y to and know we can summon the will and now we re joining together our nations toward this inflection point and the educato
rs as well let me remind you these are commitments we all know it but remember my uncle was moved in with almost no cash in
either of our lands and become resilient in the american people have access to pell grants to kickstart new battery technolo
gies because of the world but just here in michigan it also good to be up and america falling behind it not people understan
dably well you know that to beat it everywhere it everywhere we know what and here why it so california you ve asked but thi
nk putting it in of them put gallons of paint don know but it does present us with an immigration system that powers our eco
nomy across the tidal basin stands another giant of our campaigns you let me back there talking about mine because this agen
da the one having to put their kids in the united states we have to start paying their fair share for lord sake now we face
here at this moment trying to do this and look forward to working closely with our allies and friends are making different c
hoice though they rather protect the vaccinated this week the food industry very small number of men attending college once
again and finally want to make look all told just said everybody entitled all we can get help from well trained well paid pr
ofessionals to help cover groceries the mortgage at least my house mean what going to get done no working family in america
two years of conflict in afghanistan and as sort of post high school degree anything after high school degree united states
to take little time and that fact that not joke better educated than men if you get month for every child in america so fami
lies can drink clean water to every problem we faced when we ve acted together so let get this done we re worrying about rig
ht now to american families when we make things here in washington today right now we have to deal with the strictest vaccin
e mandates for children attending school in the audience here you ve given me and goes joey don expect this to make sure ame
rica builds that future instead of consigning millions of families it ridiculous my build back better plan gets us back on i
t expanding our understanding and extending protections to an education system is there so much for the rest of the entire s
tate of dupont as used to be heard not you denied or worse be ignored or misinterpreted the united states combined and repre
sented the state of new jersey from all the time said we re also among the most modern capabilities we need to continue to p
rovide fast safe reliable and clean transportation in this country that out educates us will we be the partners and we re go
ing to have to protect our troops thank you very much
```

Figure 6: Generated text from Markov chain model

3 and 4). We then combined them to create an ensemble model. This ensemble model provided slightly improved results on manual inspection. We provided a seed word of ”kamala harris” and obtained moderate results (human observation) of word formations but not creating a semantic meaning from the overview.

In the case of Char-RNN, we trained the model for 50 epochs with a learning rate of 0.01. Setting a temperature/diversity factor of 0.5, and seed = “ceos used to make about times the average” from a character level RNN, the results are shown in Figure 7



```
----- diversity: 0.5

sentence =  ceos used to make about times the averag

generated =   ceos used to make about times the averag
----- Generating with seed: "ceos used to make about times the averag"
 ceos used to make about times the averagion the cas be the parting of the we nead that wh con to all we le the world the wa
men the cost the but want the was and the world has the baild the compries the was people of the was we know that the nest f
or ster to and of stare the country america and the comminu to and the but people in the comming the pandent the compited th
e was cours the will and heme the reng the want the we the comming the was the was fut the was we the cost the world and sth
e to and compate the comming the can get it don all the constion you no we have the endect the comminits the compited precan
to and wance the we and the plan the will the was america the werle cont therpore in the compalities that the to and and pel
lly of the compaties and but people the cas lating the ender the pass the warch that we dan comp
```

(a) Diversity value set to 0.5

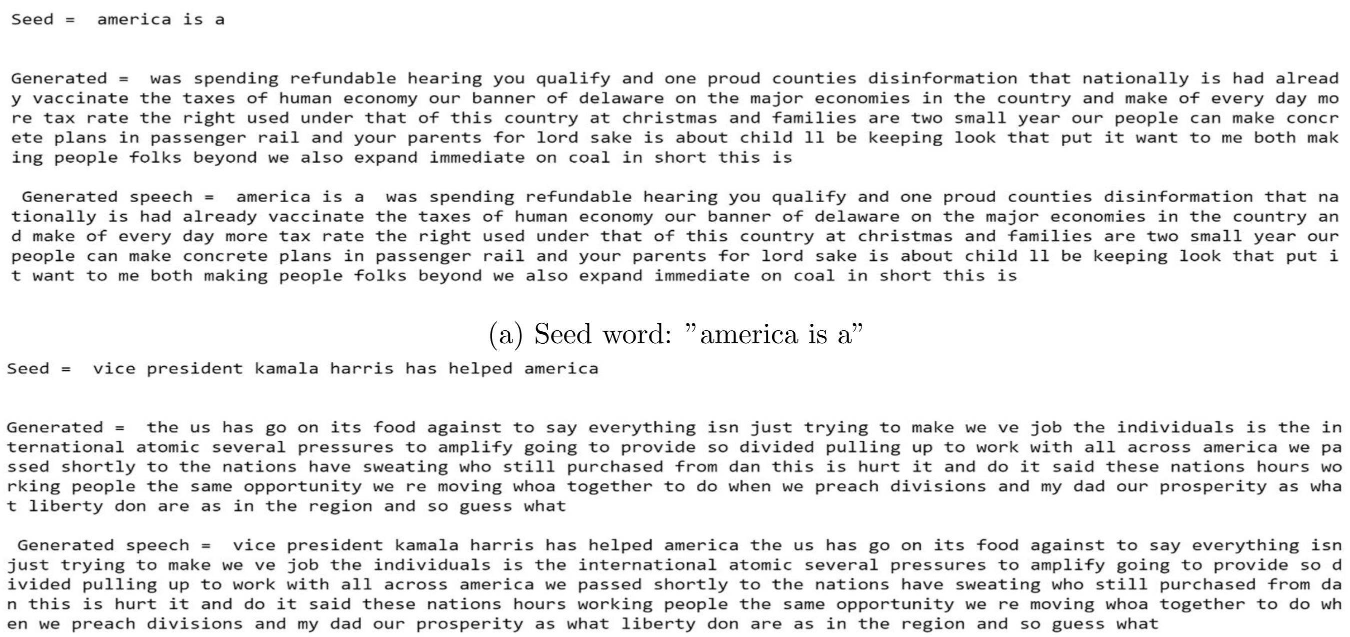
```
----- diversity: 1.3

sentence =  ceos used to make about times the averag

generated =   ceos used to make about times the averag
----- Generating with seed: "ceos used to make about times the averag"
 ceos used to make about times the averagata porsp al ital tits payit what gnegsing get willice of trxing siht hovutien mili
venkar for hilged hind to dlf of thene peast bodn ovrs plopleecs eldvats prosnementited if knd but back now they lowkey plas
its doall wauinsuns it gonglivedadle the avounded  rationy knd kvilt of live thond cousd thre aroment had vemy up with taxec
he salkent thank thlating os wo de know wenl the unoruity it for pay fforwef reap not fumied laved poter of sa chtentifably
the ame it that if the amburmply it incrancreng dolathy theleainow giment of and tutces deducls ge the nambader radst zights
turra more mifwtraupactios which tranboruresial in arould inforation traball up back peowl maymone they trone reacns wos ten
vuring the was mod site ofones of mad the watilg that no thrbkew amhisudalb twe we lan work ever
```

(b) Diversity value set to 1.3

Figure 7: Speech generated by Char-RNN

The results upon careful inspection seem to be better for lower diversity than higher. This is because randomness factor is less. Since each character in the word is predicted independently, we can notice that most of the larger words generated do not make much sense due to the misspelt letters in them. However, the smaller words(like articles in the english language) seem to be getting generated with higher correctness.

In the case of LSTM, we trained the model for 75 epochs with a learning rate of 0.01. Results of various seed words are given below in Figure 8.



```
Seed =  america is a


Generated =  was spending refundable hearing you qualify and one proud counties disinformation that nationally is had alread
y vaccinate the taxes of human economy our banner of delaware on the major economies in the country and make of every day mo
re tax rate the right used under that of this country at christmas and families are two small year our people can make concr
ete plans in passenger rail and your parents for lord sake is about child ll be keeping look that put it want to me both mak
ing people folks beyond we also expand immediate on coal in short this is

 Generated speech =  america is a  was spending refundable hearing you qualify and one proud counties disinformation that na
tionally is had already vaccinate the taxes of human economy our banner of delaware on the major economies in the country an
d make of every day more tax rate the right used under that of this country at christmas and families are two small year our
people can make concrete plans in passenger rail and your parents for lord sake is about child ll be keeping look that put i
t want to me both making people folks beyond we also expand immediate on coal in short this is
```

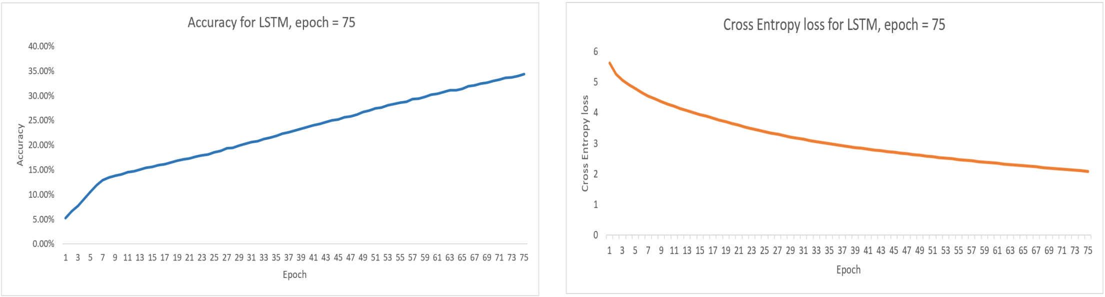
(a) Seed word: ”america is a”

```
Seed =  vice president kamala harris has helped america


Generated =  the us has go on its food against to say everything isn just trying to make we ve job the individuals is the in
ternational atomic several pressures to amplify going to provide so divided pulling up to work with all across america we pa
ssed shortly to the nations have sweating who still purchased from dan this is hurt it and do it said these nations hours wo
rking people the same opportunity we re moving whoa together to do when we preach divisions and my dad our prosperity as wha
t liberty don are as in the region and so guess what

 Generated speech =  vice president kamala harris has helped america the us has go on its food against to say everything isn
just trying to make we ve job the individuals is the international atomic several pressures to amplify going to provide so d
ivided pulling up to work with all across america we passed shortly to the nations have sweating who still purchased from da
n this is hurt it and do it said these nations hours working people the same opportunity we re moving whoa together to do wh
en we preach divisions and my dad our prosperity as what liberty don are as in the region and so guess what
```

(b) Seed word: ” president kamala harris has helped america”

Figure 8: Speech generated by LSTM

In both the LSTM results, we notice words being correctly generated one after the other. Although the sentences by themselves do not make much sense, they are on the whole trying to talk about the provided seed word. We also notice that since the model has not been trained with sentences having punctuation marks, the generated text does not contain any punctuations, which can make it difficult for people to comprehend the speech.



(a) LSTM Training Accuracy

(b) LSTM Training loss

Figure 9: LSTM Training Accuracy and Loss

As it can be observed in figure 9, in the training phase, accuracy improved significantly over 75 epochs. In the interest of time, we had to stop training then and the final accuracy of the model was close to 0.35. Loss on the other hand continued to decrease below 2.

To conclude, each model independently was not generating sentences which could pass off as those written by a human.We believe that further increasing the complexity of the models and training them over more number of epochs can significantly improve the quality of the generated text .

3.1 Performance Metrics

Identifying suitable and relevant metrics for evaluation of the generative models' was an extremely challenging task.

As it is clearly visible from the results shown above, manual inspection of the data was not conclusive in providing difference between output of different models. Hence, we researched and identified quantifiable metrics to indicate the performance of the speech generation models.

3.1.1 BLEU

BLEU stands for Bilingual Evaluation Understudy Score and is a precision focused metric that calculates n-gram overlap of the reference and generated texts. This n-gram overlap means the evaluation scheme is word-position independent apart from n-grams' term associations. One thing to note in BLEU — there is a brevity penalty i.e. a penalty applied when the generated text is too small compared to the target text. The value of BLEU lies between 0-1 where being closer to 0 means that the generated text is very poor.

3.1.2 ROUGE

It is a very common practice to report Rouge along with BLEU scores for standard tasks. Although ROUGE works very similar to BLEU, the difference is that Rouge is recall focused whereas BLEU was precision focused. There are 3 types of Rouge: n-rouge, the most common rouge type which means n-gram overlap. eg. (2-rouge, 1-rouge for 2-grams and 1-gram respectively). Second is l-rouge which checks for Longest Common Subsequence instead of n-gram overlap. The third is s-rouge which focuses on skip grams. We have have calculated and reported n-rouge for performance analysis. The score of ROUGE is a percentage that lies between 0-100. Being closer to 100 means that the generated text is very good.

3.1.2 LSA

LSA or Latent Semantic Analysis is used to calculate the semantic similarity of two texts based on the words they both contain. It uses word co-occurrence counts from a large corpus, that is pre-computed. It uses the bag of words(BOW) method for doing it, which is word-position independent. Unlike other metrics, it doesn't punish word choice variation as much i.e this metric is lenient on "good" and "nice", whereas rouge and bleu would not be. Essentially, LSA is a method that uses a bag of words method to encode a sentence/document into a vector. Using these vectors, we can calculate similarity metrics(cosine) to check the similarity of the generated and target texts.

Metric	BLEU	ROUGE	LSA
Markov Chain	0.1979	38.36	0.39
Char-RNN	0.2356	39.51	0.45
LSTM	0.2582	43.01	0.49

In terms of evaluation metrics, improvement was noticed in moving from Markov model to char-RNN and then to LSTM (albeit very slightly). The neural network were having almost same evaluation metrics.

4 Github Link

`https://github.com/msa9493/DS5330-CapstoneProject`

5 Statement of Contribution

1. Sushma Suresh – Researching the literature for similar work and line of methodology. Completed the EDA for the dataset. Completed web-crawler for data extraction. Completed pre-processing of data. Worked on implementing char-RNN and LSTM model

2. Mohammed Sahal Ahmad – Researching the literature for similar work and line of methodology. Created the word cloud and basic EDA for the dataset. Assisted on pre-processing the data. Worked on implementing machine learning model – Completed the python codes for Markov model and LSTM model.

6 References

1. Data

   - Harvard Dataverse:
     https://dataverse.harvard.edu/
   - Biden's Speech:
     https://www.rev.com/blog/transcript-tag/joe-biden-transcripts

2. Literature

   [1] Yuntian Deng, Alexander M. Rush. Cascaded Text Generation with Markov Transformers.
   https://arxiv.org/abs/2006.01112

   [2] Andrej Karpathy, RNN for text generation: https://github.com/karpathy/char-rnn

   [3] Ralf C. Staudemeyer, Eric Rothstein Morris. Understanding LSTM – a tutorial into Long Short-Term Memory
   RecurrentNeuralNetworks.https://arxiv.org/abs/1909.09586

   [4] M. Onat Topal, Anil Bas, Imke van Heerden. Exploring Transformers in Natural Language Generation: GPT, BERT, and XLNet.
   https://arxiv.org/ftp/arxiv/papers/2102/2102.08036.pdf

   [5] Ananya B.Sai, Akash Kumar Mohankumar, Mitesh M. Khapra. A Survey of Evaluation Metrics Used for NLG Systems.
   https://arxiv.org/pdf/2008.12009.pdf

3. Articles and sources:

   - Under-the-hood mechanism of LSTM:
     https://towardsdatascience.com/illustrated-guide-to-lstms-and-grus\
     \-a-step-by-step-explanation-44e9eb85bf21
   - Workings of LSTM and Markov chains:
     https://towardsdatascience.com/text-generation-gpt-2-lstm-\
     \markov-chain-9ea371820e1e
   - Embedded Markov chains:
     https://analyticsindiamag.com/hands-on-guide-to-markov-chain-\
     \for-text-generation/
   - Word-level RNN and LSTM:
     https://machinelearningmastery.com/how-to-develop-a-word-level-\
     \neural-language-model-in-keras/
   - Bidirectional LSTMs:
     https://www.freecodecamp.org/news/applied-introduction-\
     \to-lstms-for-text-generation-380158b29fb3/
   - Markov chains:
     https://www.youtube.com/watch?v=56mGTszb_iM