Assignment report on

"Script Generation and Comparison using RNN with LSTM and Bi-Directional RNN with LSTM"

Submitted for Topics in Deep Learning during 7th semester of

Bachelor of Technology
in
Computer Science & Engineering

Submitted by :

01FB16ECS006 : Aayushi Sanghi
01FB16ECS199 : Mayank Rao
01FB16ECS301 : Rhiya Ramesh

August – December 2019

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

PES UNIVERSITY
(Established under Karnataka Act No. 16 of 2013)
100ft Ring Road, Bengaluru – 560 085, Karnataka, India

Problem Statement

Generation of new scripts, while trying to preserve semantics (context) and syntax, for existing novels. (Harry Potter and the Philosopher's Stone has been used to train the models described below).

Dataset

All chapters of Harry Potter and the Philosopher's Stone.

Approach Taken

In this project, we have considered two approaches and have compared them against each other.

1. Unidirectional character-based RNN with LSTM cells:

   The main task of the character-level language model is to predict the next character given all previous characters in a sequence of data, i.e. generates text character by character (based on a step value defined by the user).
   In this approach, each character used in the dataset is encoded into an integer. Encoding the characters as integers makes it possible to use as input in the network. Since the network is working with individual characters, it's similar to a classification problem in which we are trying to predict the next character from the previous text. The number of classes to pick from would be the total number of unique characters used in the dataset.
   Training is done in batches which are multiple sequences of some desired number of sequence steps (a model parameter).
   While training, the input array is shifted by one and fed as target array as that's what it is supposed to output in the best case.
   Normal RNNs have the issue of gradients either exploding or disappearing. LSTMs fix the disappearance problem, but the gradients can still grow without bound. To fix this, we can clip the gradients above some threshold. That is, if a gradient is larger than that threshold, we set it to the threshold. This will ensure the gradients never grow overly large. Then we use an AdamOptimizer for the learning step.

Once the network is trained, we'll use it to generate new text. The idea is that we pass in a character, then the network will predict the next character. We can use the new one, to predict the next one. And we keep doing this to generate all-new text. We also included some functionality to prime the network with some text by passing in a string and building up a state from that.

2. Bi-directional RNN with LSTM and Doc2Vec:
   The principle of BRNN is to split the neurons of a regular RNN into two directions, one for positive time direction (forward states), and another for negative time direction (backward states). Those two states' outputs are not connected to inputs of the opposite direction states.
   In this case, unlike the earlier case, we use word-based RNN. Since the context is not captured in character-based RNNs, we use word-based RNNs with the hope of generating semantically correct sentences.

   The first bidirectional LSTM is used to generate new sentences, word by word.
   Using this neural network, we generate several sentences, which are candidates to be the next phrase of the existing text.
   Then, we use a second Bi-Directional LSTM to select the best candidate between all these sentences.
   How this is done, is that the second bi-directional LSTM takes in 15 vectors to predict the "best sentence" - which is also a vector.
   We then use cosine similarity to compare all candidates with the "best sentence" and thus the "select_next_phrase" function selects the candidate which is closest to the "best sentence."

Challenges:

1. **Batch Size and GPU utilization :** Initially, while performing character level modelling, we encountered several problems. Firstly, the batch sizes being too small would default to the CPU over the tensorflow-gpu. To combat this, batch size was raised in exponents of 2, until a satisfactory GPU utilization was achieved. However, the cost in increasing batch size was lower quality outputs.

2. **Spelling errors :** Due to the nature of character-level model, each word is built individually as per the neural network's training. These spelling errors would ideally reduce over the span of training, however even at over a hundred epochs, spelling errors were rife in the output generated text. This was remedied simply by using a custom made Minimum Edit Distance formula. The formula works over the entire vocabulary of the book and trims words that occur less than 10% of the time. This ensures that words don't get corrected to their outliers very often, as well as speeds up calculation due Minimum Edit being an O(n^2) algorithm that performs pairwise search amongst every element in the vocabulary with every wrong word encountered .
    a. This formula ranks the possible word output in order of :
        i. **Similar string length :** This ensures that Minimum edit distance remains as small as possible. Ideal strings are searched for within +2 length to -2 length. Meaning a string can either grow by 2 characters, or reduce in size by 2 characters.
        ii. **Edit Distance :** The edit distance itself. O(n^2) algorithm that returns the best possible word that can be created by inserting, updating or deleting characters. (Assumes cost of all operations is 1)
        iii. **Jaccard Similarity :** Defined as

$$J_\mu(A, B) = \frac{\mu(A \cap B)}{\mu(A \cup B)},$$

Where A is the word being checked and B is a word to be corrected. All the words are ranked on all 3 parameters in that specific order, and the returned word is overwritten accordingly, if it is deemed to be an error.

3. **Word Level RNN :** An alternative that was considered was a word level RNN. This would solve the issue of spelling errors as previously defined, and the code was hence modified to encode words instead of characters, following the same logic as previously defined.

   This problem would be better solved using Doc2Vec and a second bi-directional RNN, which was the second architecture we implemented and tested.

4. **RNN Layers :** Due to the loss in quality from having to increase batch sizes for efficient training, we decided to increase the number of layers to hopefully capture a deeper understanding with the network. However, as was expected, deeper layers do not work properly with RNNs and the text ended up being a garbled mess.

5. **Initially started with script generation:** Our initial problem statement was generation of scripts using these exact same architectures, however since scripts require a lot of structure in the way they are typed out (eg: Monica: How was your day? Rachel: It was exhausting!) and because they require a lot more emoting, the results we were getting weren't too satisfactory. After training several models and getting much better results, we decided to try this on books, as we expected them to perform better. According to our expectations, they did, and we decided to go ahead with training models on books to see how they would perform.

Architecture of Models:

1) **Architecture of the character-based RNN with LSTM cells:**

After experimenting with different numbers of layers, we concluded that 2 layers of LSTM cells work best for the network. Each layer has 128 LSTM cell units.

It is a seq-2-seq model which means that given a sequence of length "num_steps", the target is the same sequence shifted by 1.

We have used dropout in our layers, with a keep probability of 0.6. This has been done to prevent overfitting and normalizing our results.

The hyperparameters used for this network are:

- Batch size: 128 sequences
- Number of steps: 50
- Learning rate: 0.001
- LSTM_size: 128 cells

Loss for this model: 1.18

2) **Architecture of Bi-directional RNN with LSTM and Doc2Vec setup:**

1) Architecture of First Bi-Directional RNN:

Hyperparameters used for this network are:

- RNN size: 256 cells
- Sequence length: 30 words
- Learning rate: 0.001

Loss for bi-directional LSTM 1: 4.0527

2) Architecture of Second Bi-Directional RNN:

- RNN size: 512 cells
- Sequence length: 15 vectors

Loss for bi-directional LSTM 2: 0.0535

Results:

1. Character based RNN with LSTM:

- 200 Epochs without correction with learning rate = 0.001 with batch size = 128

```
THE MIRROR OF ERISED, then there aren't tow the thirror fle toer it a stald a the toll out of a
sitter then to the saming. The chearsard as Hagrid whispers, a the hall tried the cross from thems
was show on him of that stind here. However he'd he heard and a fow their spacker for he'd his broom
at the hat tabked and sang anything out the stuntes to strinking in the corridor the finde, but a
but wouldn't say he was to steam to head the glicking and too studden to take the drowes of them. He
caned his head out as his he'd as that was a to his hard. "Seen and would and year. All on thooge
around to head the store - what have gos the start, all soed your feels at the torrow with, his had
tell you and they said, Hogwarts. A shall one all sore off of the trounnt out as anyone, bothing it
with the class that seanted. "He whole you hand the floor and were to the comming through that
thought a trances and then are at themed at the troll's anting themed, with the forest of the
cander. He hadn't been the they comeshing the thingss to stoppor thas. It was all toward the talles,
were the chest to hand him. "He could sele you the grounds and he asked as he went interthing to the
feots of a lott and whinged and weally sited. "Thought," said Ron. "I've to so me a told and a broom,
 him, and so seater to think, Hogwarts, Harry's beand ans allowad." He stored it hand on it had back
the second and heard the going two they wanting his books it on his hand ot had hears as a feeling
the thould have taken. "Wall the corridor, around to the say of their find, they cent on the
granster wanted as a the wall, his stell, what him any alray. Harry had at that a fimpo him as he
was how hopesing, the colling watched the teachers as Harry a ghund at his. "And his broomstors?" "I
```

- 2000 Epochs without correction with learning rate = 0.001 with batch size = 128

```
Invisibility Cloak on top of the towers and stants of him to his hard, smilling in the fire, which
was straight than him. "I'm brothing the Stout open onto" that he wasn't blungainstates and his
hands in because as the ghust something. Harry suddenly look ut the four before the sort of the
field with his sight. "We with your more to the stairs." Hermione was a lumpy, which was something
to see the trainsofriss, buttles. Harry had seen a ballood, and study to sige. "What we ang a bit
the fire. They walking allowed to trick a tall." He saw, but it on the crate and holring. They
seemed to have leaded at this. "A womded all was the tark. Harry was going to catch the book, his
teaches with blec to thoughther than your first headly. "And then what all you dad, I're never heard
it around." "Wing inas, they have been say toward the forestainst the groun and becoused it." "The
feals on the saye to see it will get into the cloak and seven yearly was going to get a mount. He
couldn't have hundreds. "I suppose to she to hell?" "And I'd see on, they'd heard Slytherin. He had
been back into the directions of the trophed at the front were a teeroul and that trying to told the
troll was too stratgedly in a smanly wincer hurried spickstion back and left the stell, this all the
face by the second while with the tarchear. Hermione and Ron stick and then, because Ron already
care and have to stait floaded in the class. "Wertord you the books walking in," said Hermione.
"He's going to see a few of his feet!" He desped at it. The hours, and so the class what the troll
colled a lole with onling the crowd before in his said of the top of one of the fould was to seeved
his legg. Harry had started to inside. "It!" said Professor McGonagall they come another his head,
suddenly they had been looked about the teachers. It was the bush. "Wandering," said Ron, "He was
stinling it with a bott of the deskly to him. "You're never been in a thon's beave and a stat ar in
it. Hermione had so dasten. At the forest were about it, and he was flicked on the stats flathing
and took his field the cold, because Hagrid sags off. "I he's not been." "And any if," said Harry,
but he was so to be to hins the firs from the fell their while a long bit and spokestair. Hagrid
```

- 2000 Epochs with correction with learning rate = 0.001 with batch size = 128

```
Invisibility Cloak on top of the toward and stands of him to his hard, thinking in the fire, which
was straight than him. "I'm breaking the snout open onto" that he wasn't particularly and his hands
in because as the ghost something. Harry suddenly look ut the four before the sort of the field with
his sight. "We with your more to the stairs." Hermione was a lucky, which was something to see the
particularly, between. Harry had seen a allowed, and study to side. "What we and a bit the fire.
They walking allowed to thick a tall." He saw, but it on the crate and holding. They seemed to have
leaned at this. "A wonder all was the dark. Harry was going to catch the book, his teacher with flew
to themselves than your first nearly. "And then what all you did, I are never heard it around."
"fang into, they have been say toward the interesting the green and realized it." "The heads on the
safe to see it will get into the cloak and seven nearly was going to get a found. He couldn't have
hundreds. "I suppose to she to fell?" "And I'd see on, they'd heard Slytherin. He had been back into
the reflection of the dropped at the front were a several and that trying to told the troll was too
restricted in a really wonder hurried reflection back and left the smell, this all the face by the
second while with the teachers. Hermione and Ron stuck and then, because Ron already came and have
to start pleased in the class. "nervous you the books walking in," said Hermione. "He's going to see
a few of his feet!" He gasped at it. The doors, and so the class what the troll called a hole with
during the crowd before in his said of the top of one of the could was to seemed his legs. Harry had
started to inside. "It!" said Professor McGonagall they come another his head, suddenly they had
been looked about the teachers. It was the busy. "Wandering," said Ron, "He was standing it with a
both of the nearly to him. "You are never been in a than's leave and a seat ar in it. Hermione had
so listen. At the forest were about it, and he was knocked on the stand floating and took his field
the cold, because Hagrid ears off. "I he's not been." "And any if," said Harry, but he was so to be
to find the fire from the fell their while a long bit and understand. Hagrid said on the Gryffindor
```

As we trained the LSTM for longer (2000 epochs), we noticed that the results got a lot better. However on increasing the number of layers, we found that the network got confused and produced worse results.

- 250 Epochs without correction with learning rate = 0.1 with batch size=256

```
Invisibility Cloak on top of the tower," shatt he'll the would to he as the wand this wonder think
that all, sat, all ho hee shant thit thrite we whed with to home. The sas said, he said. "Tow he at
a had wann. He the thould and was was second the a antrets, the has the seal as somethed as than at
sees was wording tham the a tarrad he any a his, alound they sanding he's the so and seorst and sat
as he that was, he and are wonderste thousing. He all thas and thougut to something. Nen to and was
stints he seone was, seoled was hee a tous trast, weat she sand that a wering thought, he term was,
honas, whann to the thicks, thould the a steal, at of ho stoting to the thontoten at shook when one,
" Ron his and and what and to the and was his thritch's alled, we and as hook alet the toll to take
the thole. "Thank and him think wontay the toll thoute at the ter all seortastn, we'd his hee seole
to that as of steast on stame the a hull star when on he than shrore shooked hudding out there to
when one he's at, a stead an he the shade to and sand sat as said, they wome his, all thing a shith
of the tholl hus and were with he shatting though sanded we're head, her an' he to tere, tidice was,
and the trost the hith at his that that a sand wesit have at this. Hermione to we stither the whore
stared. It and here to heltense sealed, hoot, a womped. Hall seal tecn a see them stis they the heal
he're that he and as had to all as that's a hat the sall harred. "Nothers him to that went on the
sascalk at the tall a a theed, and she so and we sat wass the strested there, think to his hols well
shant sead the telled a head," said Tham the his a steet teared to the wasing almost heard they the
selte strace they's seally. He stood him worted, stoll and that he to the with." And white them so
them, that and to them what. He to and wannered. It a as at he well to was was too thould wompaced
```

- 250 Epochs with correction with learning rate = 0.1 with batch size=256

```
Invisibility Cloak on top of the tower," start he all the would to he as the wand this wonder think
that all, sat, all ho lee spent that during we when with to home. The has said, he said. "how he at
a had wand. He the should and was was second the a another, the has the seat as together as than at
seem was working team the a hagrid he any a his, around they wanting he's the so and secret and sat
as he that was, he and are wondering thousand. He all than and thought to something. new to and was
stands he stone was, seized was lee a does train, seat she hand that a during thought, he team was,
hoops, learn to the thanks, should the a steal, at of ho staring to the forgotten at shook when one,
" Ron his and and what and to the and was his through's asked, we and as book ages the tall to take
the those. "think and him think wonder the tall though at the ter all important, we'd his lee score
to that as of please on snape the a full stay when on he than before shouted holding out there to
when one he's at, a steal an he the snape to and hand sat as said, they come his, all thing a which
of the troll has and were with he standing though wanted we are head, her an' he to here, hiding was,
 and the front the high at his that that a hand visit have at this. Hermione to we another the whole
stared. It and here to hermione seized, foot, a jumped. Hall seat dean a see them said they the head
he are that he and as had to all as that's a hat the ball hagrid. "another him to that went on the
marched at the tall a a their, and she so and we sat miss the screamed there, think to his hole well
spent head the yelled a head," said team the his a sleep leaned to the waving almost heard they the
crate stared they's really. He stood him wanted, still and that he to the with." And white them so
them, that and to them what. He to and wondered. It a as at he well to was was too should wondered
```

2. Bi - directional word based RNN-LSTM with Doc2Vec:

Training of the first Bi-RNN model:

```
Layer (type)                 Output Shape              Param #
=================================================================
bidirectional_1 (Bidirection (None, 512)               8171520
_________________________________________________________________
dropout_1 (Dropout)          (None, 512)               0
_________________________________________________________________
dense_1 (Dense)              (None, 3733)              1915029
_________________________________________________________________
activation_1 (Activation)    (None, 3733)              0
=================================================================
Total params: 10,086,549
Trainable params: 10,086,549
Non-trainable params: 0
_________________________________________________________________
WARNING:tensorflow:From c:\users\mayan\appdata\local\programs\python\python37\lib\site-packages\keras\backend\tensorflow_backen
d.py:422: The name tf.global_variables is deprecated. Please use tf.compat.v1.global_variables instead.

Train on 40369 samples, validate on 408 samples
Epoch 1/50
40369/40369 [==============================] - 161s 4ms/step - loss: 6.3550 - categorical_accuracy: 0.0510 - val_loss: 6.0381 -
val_categorical_accuracy: 0.0833
Epoch 2/50
40369/40369 [==============================] - 155s 4ms/step - loss: 5.8264 - categorical_accuracy: 0.0875 - val_loss: 5.9484 -
val_categorical_accuracy: 0.0931
Epoch 3/50
40369/40369 [==============================] - 158s 4ms/step - loss: 5.6087 - categorical_accuracy: 0.1090 - val_loss: 5.8761 -
val_categorical_accuracy: 0.1275
Epoch 4/50
40369/40369 [==============================] - 157s 4ms/step - loss: 5.3876 - categorical_accuracy: 0.1364 - val_loss: 5.8572 -
val_categorical_accuracy: 0.1446
Epoch 5/50
40369/40369 [==============================] - 156s 4ms/step - loss: 5.1652 - categorical_accuracy: 0.1572 - val_loss: 5.7989 -
val_categorical_accuracy: 0.1618
Epoch 6/50
40369/40369 [==============================] - 157s 4ms/step - loss: 4.9518 - categorical_accuracy: 0.1766 - val_loss: 5.7743 -
val_categorical_accuracy: 0.1691
Epoch 7/50
```

Output of the first Bi-Directional RNN: (Without usage of Doc2Vec to enhance output)

```
. i'm no, said ron, hagrid's your sorcerer's stone, he'dn't a the. hagrid, the yelled, but he's must be remembrall. he asked ro
n, but what's he's than, said wood. i'm to the remembrall that, you've got all for the team. don't want to you, said hermione.
i've got the see you, said ron. you're too not to your i think i'm in the team, said ron. he said. oh, you're not me to you've
got to get. he is it. i'm i know you? well, said ron, harry's time to have to be of the back. i'm' - i've got the door, said ro
n, and harry had had to percy on the gryffindor, but it's the too. harry's were the head, said ron, harry at once. the new. h
e's really've got to the the dragon's. you'll have to have been a a this, and snape wasn't have the the in a the few and that w
as three of the gryffindor. i'm this, you'd be were no. ron, said ron. i'll not him, said ron. i'm you get a it. i'm me - you'r
e i've got to be i'm, said ron. i'm your you, said ron. i'mn't come, said harry. i'm i'm' you, said ron, i'm you, you're going
to be the house the first years in the mirror of and then i'll be the. i don't want to a know, said ron. i'm, said ron. how to
you, you've got to get a the away and year, that's the midnight your high, up? said ron, you've got the head, said ron, the gry
ffindor, said ron. i'm about you, said ron, harry told ron. what are you? said, hagrid's us the you're an''. said, said harry.
ron, said ron, what i'm the other, said hagrid. you've got to very going to tell what about it, you'll be you? i'll get me for
you, said harry, harry's be a his own. you've got to hear you will be you've got about gryffindor, said ron, ron as they tree.
then, but the harry, was, but i've got to it, then, but they couldn't for? said, said harry. the house on the the team, , but'
harry's got to the and. so you could have have to be charlie to be of the brooms. i'm trying to be going to be the high, harry
and harry looked him. he couldn't a a you, you've got to a - of the head, and ron had been this. he hadn't think they were, the
the house. i'm he didn't see you, said hermione, ron was, but he didn't be him. that was you, her from the dog, but they were h
ave been a was nimbus two thousand. ron, ron, i'll be stupid, said harry, but ron was in the the floor, and and the - quidditch
and a just and a, the eyes of the was broomstick, that was as they his long as they were no with a hands, and they had a knocke
d. the went to a time to have a the between. the whole of the troll, the floor. you've got to see you, you're a'd seen, harry,
said ron. he's got the the school, said harry. i don't a a get it. you don't see you, said harry. the nimbus two thousand. pott
er, said ron. i'm, but you're a you around, i know you're. you're a, said dumbledore. harry had about the found a which were th
ey looked at up. harry's other. i don't see you, you're? said snape. i'm i've got to be seeker, i think that was be can you? i
don't you you? said hermione. i'm told you, said harry. i've got to see
```

Training of the second Bi-RNN model (Post Doc2Vec improvisation):

```
Train on 2340 samples, validate on 260 samples
Epoch 1/40
2340/2340 [==============================] - 5s 2ms/step - loss: 0.0622 - acc: 0.0256 - val_loss: 0.0531 - val_acc: 0.0731
Epoch 2/40
2340/2340 [==============================] - 4s 2ms/step - loss: 0.0573 - acc: 0.0603 - val_loss: 0.0527 - val_acc: 0.0731
Epoch 3/40
2340/2340 [==============================] - 4s 2ms/step - loss: 0.0563 - acc: 0.0624 - val_loss: 0.0525 - val_acc: 0.0615
Epoch 4/40
2340/2340 [==============================] - 4s 2ms/step - loss: 0.0557 - acc: 0.0637 - val_loss: 0.0524 - val_acc: 0.0577
Epoch 5/40
2340/2340 [==============================] - 4s 2ms/step - loss: 0.0553 - acc: 0.0607 - val_loss: 0.0523 - val_acc: 0.0538

Epoch 00005: saving model to models\my_model_sequence_lstm.05.hdf5
Epoch 6/40
2340/2340 [==============================] - 4s 2ms/step - loss: 0.0550 - acc: 0.0598 - val_loss: 0.0523 - val_acc: 0.0538
Epoch 7/40
2340/2340 [==============================] - 4s 2ms/step - loss: 0.0547 - acc: 0.0573 - val_loss: 0.0523 - val_acc: 0.0538
Epoch 8/40
2340/2340 [==============================] - 4s 2ms/step - loss: 0.0544 - acc: 0.0624 - val_loss: 0.0522 - val_acc: 0.0538
Epoch 9/40
2340/2340 [==============================] - 4s 2ms/step - loss: 0.0542 - acc: 0.0624 - val_loss: 0.0522 - val_acc: 0.0500
Epoch 10/40
2340/2340 [==============================] - 4s 2ms/step - loss: 0.0540 - acc: 0.0620 - val_loss: 0.0522 - val_acc: 0.0500

Epoch 00010: saving model to models\my_model_sequence_lstm.10.hdf5
Epoch 11/40
2340/2340 [==============================] - 4s 2ms/step - loss: 0.0537 - acc: 0.0607 - val_loss: 0.0522 - val_acc: 0.0577
Epoch 12/40
2340/2340 [==============================] - 4s 2ms/step - loss: 0.0535 - acc: 0.0607 - val_loss: 0.0522 - val_acc: 0.0577
```

Conclusion:

After analyzing both results, we can safely conclude that LSTMs cannot be used to generate very semantically correct text with the limited data that we were able to train on, but they do generate good syntactically correct text even with limited data - for both character level RNN and word-level bi-directional RNN with Doc2Vec. While the character level model does well, it is slightly lax on spellings, because we predict output sequences character by character, and therefore the output needed to go through our spell-correction module. Outside of this, we were impressed with the output it produced on limited data, and it was able to match up with our more complex second architecture.

However, a word-level RNN doesn’t require spell checking as it outputs words from the dictionary. The world level RNN however does require some amount of variation in words because of this (the dictionary being words and not characters), which is taken care of by using a "temperature" variable. Higher the temperature is, more the variation in words picked by the "sampling" function - which is used to generate the output. This means that the words with the highest probability won't always be picked (when the temperature is slightly high), and sometimes words with slightly lower probabilities can also get picked for the next output. A temperature of 0.34 worked well for us as there was a good amount of variation in words which did make some amount of sense. As we reduced temperature, the output became way too repetitive (Because we assume one set of words always had the maximum probability of being the next output, because of their sheer count/occurrence in the dataset. In our case, the phrase ' "I'm you", said Ron ' became slightly repetitive amongst a few others, with a low temperature and hence low variation. Increasing the temperature parameter to 0.34 fixed this for us). When we increased temperature beyond 0.34, the output became way too random and hence we stuck with 0.34.

The character level RNN trained a lot faster and gave similar results to the bi-directional RNN with Doc2Vec. The second architectural setup took longer because we were essentially training three models. Since we were dealing with sentences in the Doc2Vec - creation of sentence batches, and then conversion of these sentences into vectors, was very time consuming for a large text.

We do believe much better results can be achieved by increasing the size of RNN and training the Doc2Vec model with a very large amount of data. This is because since the dictionary of the Doc2Vec model is "words" and not a set number of "characters", it would require a lot more additional data to be able to bring in a variation in the output it generates and capture more context. For a character-level RNN, the dictionary is a set number of characters and therefore the classification problem (of the next character to predict) is a lot easier too - as you are working with the biggest possible dictionary already.

This (training our second bi-directional LSTMs with Doc2Vec model on a LOT of data with adequate resources) is something we plan on doing as future work so we can do a more thorough comparison of both architectures.

However, on the amount of data we were able to train on, it is preferable to use our initial model (the character level RNN) as it is able to capture the same amount of syntax and semantics, but trains faster and with much lesser effort. On much larger data, we believe the second model will do better as it will be able to have a more vast dictionary (because of a sheer increase in the number of unique words it will train on) and be able to capture a lot more variation in its outputs.