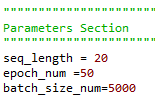
RNN Assignment

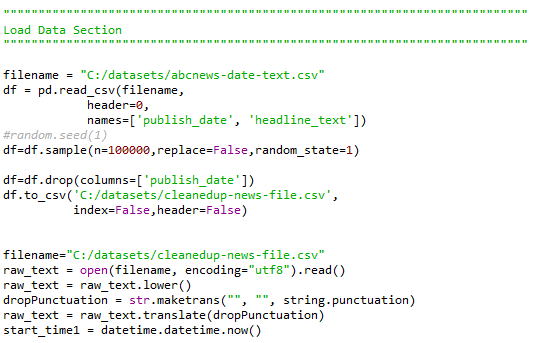
Yuke Luo

This assignment is to use recurrent neural networks to find the headlines. Every model will have 5 breakpoints to check how well the models is working and learning. I will report the result of 15-diﬀerent generated texts of length 4 to 8 words once the model is fully trained. I will do 3 experiments: Change the number of hidden units; Change the number of hidden Dense layers; and change sequence length by showing plot the training loss vs. the number of training epochs.

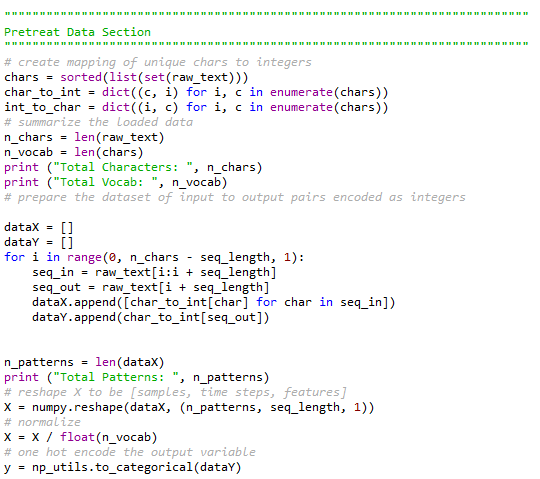
Following is my codes and explainations:



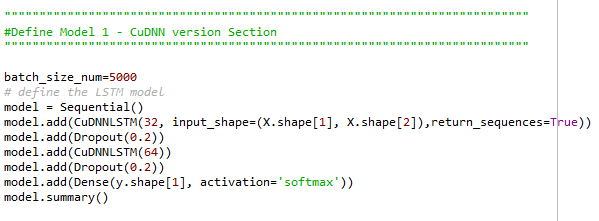
In parameters section, the sequence length is set as 20 for the basic model. The number of epochs is 50 and batch size is 5000 just to make sure that the models will have enough training. These parameters may be changed later for experiments.



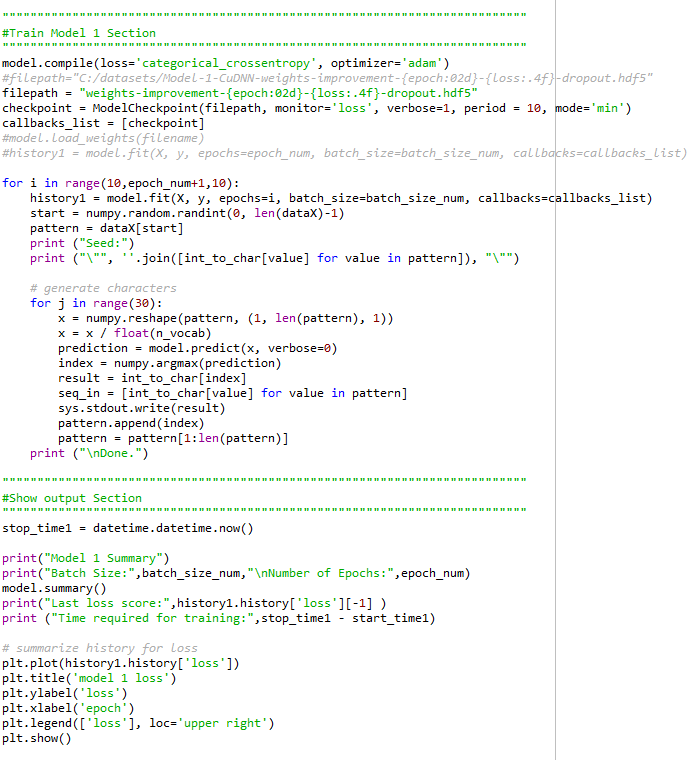
For loading data section, because the dataset was downloaded already from website and saved, it can be accessed directly. After loading data, a start time should be recorded for later use.



To pretreat the data, I first create mapping of unique characters to integers, and mapping unique integers to characters. Then, I print the length of chars and vocab to check the number of both vectors. After that is to create X and y that are needed for the model.



For the basic model, I use CuDNNLSTM in my case since my computer is able to run this model and it’s much faster than LSTM. There are 2 LSTM layers (32 and 64) with 0.2 dropout after each layer. The activation will be softmax and optimizer will be adam as usual.

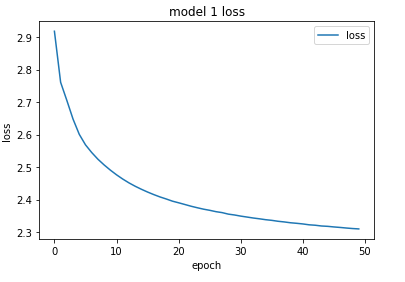


These two sections are basically just training the model and then display the summary and print the plot to see the results.

**Base Model:**

Sequence length: 20

2 LSTM layers with 32,64 hidden units each and 0.2 dropout value.



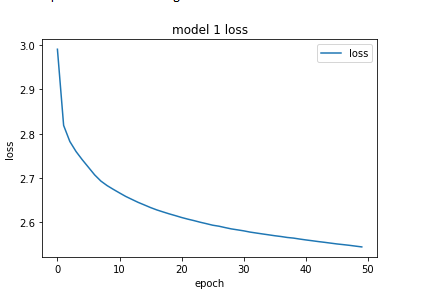
|  |  |  |
| --- | --- | --- |
| **Breakpoint** | **Seed** | **Headline** |
| 10 epochs | " ders civic hall opti " | ng to tear so tear so tear so |
| 20 epochs | " orism experts to boo " | st council to council to counc |
| 30 epochs | " hold welcome news  ta " | ys so council to council to co |
| 40 epochs | " y  report finds worki " | ng search search search search |
| 50 epochs | " ekers arrive on manu " | search search search search s |

From this chart, we can see that neither seed nor headline makes any sense. This may be caused by the model not fully trained. Therefore, model justifications are needed for the following steps.

**Halve Hidden Units Model:**

Sequence length: 20

2 LSTM layers with 16,32 hidden units each and 0.2 dropout value.



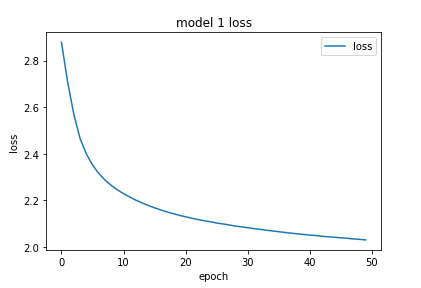
|  |  |  |
| --- | --- | --- |
| **Breakpoint** | **Seed** | **Headline** |
| 10 epochs | " care equipment  supe " | e to seseare to seseare to ses |
| 20 epochs | "  kazakhstan shuts do " | un to so so so so so so so so |
| 30 epochs | " 016 barra season dow " | n conter conter conter conter |
| 40 epochs | " ten years on  former " | to so so so so so so |
| 50 epochs | " 18 septermber 2011  r " | o contert so so so so |

Clearly, loss in this model is higher than basic model, which is not a good sign, leading us to assume this model may have the worst results. Words are repeating again and again, most of them are not even words. For the curve of loss, it’s becoming steeper than the former one. Therefore, this model should be passed.

**Double Hidden Units Model:**

Sequence length: 20

2 LSTM layers with 64,128 hidden units each and 0.2 dropout value.



|  |  |  |
| --- | --- | --- |
| **Breakpoint** | **Seed** | **Headline** |
| 10 epochs | " ons group de  fire ca " | re council to search for searc |
| 20 epochs | " rish boys behind bog " | an council search for search f |
| 30 epochs | " ccess  philippine jou " | rnalist search for search for |
| 40 epochs | " lain murder  watters " | continues for search for searc |
| 50 epochs | " rks hunt  demons focu " | s and council to search for se |

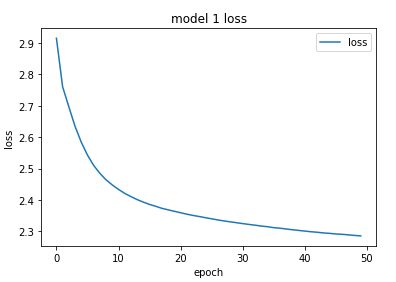
This model has an obvious decrease in loss with 50 epochs, which is good for us. And clearly the model is doing a great job in prediction. However, this model takes a lot more time than the previous two because of the doubling in hidden units.

**Additional Dense Layer Model:**

Sequence length: 20

2 LSTM layers with 32,64 hidden units each and 0.2 dropout value.

Hidden dense layer with 64 hidden units



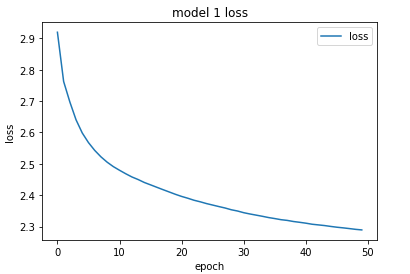
|  |  |  |
| --- | --- | --- |
| **Breakpoint** | **Seed** | **Headline** |
| 10 epochs | " ady to work with rud " | ney search search search searc |
| 20 epochs | " lice confirm woman a " | lleged to continues continues |
| 30 epochs | " for missing fisherma " | n continues country continues |
| 40 epochs | " vote in knife edge " | to continues to continues to c |
| 50 epochs | " pay  peru completes " | second second second second se |

We can tell from the graph that the loss graph suggests more training can be done. Also, the level of loss is similar to the basic model. Therefore, adding one more dense layer with 64 hidden units seems not necessary. The 50th epoch’s headline repeats second several time.

**Double Sequence Length Model:**

Sequence length: 40

2 LSTM layers with 32,64 hidden units each and 0.2 dropout value.



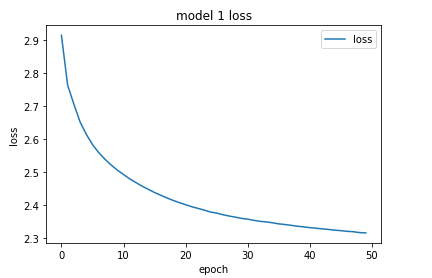
|  |  |  |
| --- | --- | --- |
| **Breakpoint** | **Seed** | **Headline** |
| 10 epochs | " the world  us intere " | st council to search council t |
| 20 epochs | " n shaped b 15t heade " | rs to search continue to searc |
| 30 epochs | " wy task force pushes " | to search continue to search |
| 40 epochs | " kambah supermarket r " | esearch council to continues c |
| 50 epochs | " to keep marine centr " | e sesearch sesearch sesearch s |

Doubling sequence length of the model returns a result of loss just similar to the basic model as well. Again, the loss decreases smoothly, but not as low as double hidden units model did. For seed and headline part, they are still not making much sense to me since headline is repeating the same words again and again. There are, however, some words in seed like “force”, “pushes”, “keep”, and so on.

**Half Sequence Length Model:**

Sequence length: 10

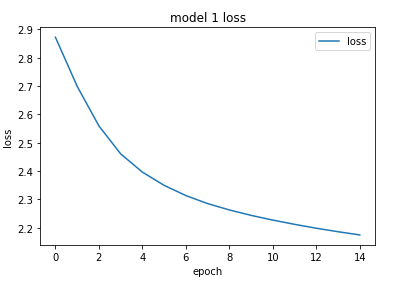
2 LSTM layers with 32,64 hidden units each and 0.2 dropout value.



|  |  |  |
| --- | --- | --- |
| **Breakpoint** | **Seed** | **Headline** |
| 10 epochs | " r  cronulla quiet one " | sesting sesting sesting sesti |
| 20 epochs | " lee lead aussie figh " | t sesting sesting sesting sest |
| 30 epochs | " es facing the wrath " | to sesting sesting sesting ses |
| 40 epochs | " flags federal charte " | rs sest to continues to contin |
| 50 epochs | " lands rejects far ri " | oe to sesting sesting sesting |

Smaller sequence doesn’t seem to have much influence on the loss or to the headline. It’s for sure no better than doubling the hidden units. One thing to notice is that the number of repeating words is much more than before, in this case, “sesting” repeats 13 times.

The last model for the 15 headlines, I chose to use Double Hidden Units Model since it has the lowest loss and makes more sense. However, I only use 15 epochs and generate seed and headline every epoch to save time. The model is fully trained.



|  |  |
| --- | --- |
| **Seed** | **Headline** |
| " fighting rethink  war " | ned of search to search to sea |
| " ce urged to address " | overtigation to search to sear |
| " odyay residents call " | stralian sestice sestice sesti |
| " era house  share mark " | inst council continues to sear |
| " ie dress crashes web " | oe to sesting sesting |
| " murders  food indust | to tear so tear so tear |
| " t qld  govt tells san " | awir aacuction coast council to |
| "  miners hitback at s " | see tor hn court over sace man |
| " g up to be one of th " | contice conmer conmer conm |
| " nson  barnaby joyce u " | se search to search to |
| " stic violence victim " | council continues to search |
| " d fiscal outlook  gc1 " | s aod roedeck to be released |
| " ed over nightclub in " | soarkes conti conti conti |
| " e  philippines electi " | nean sear coune coasirnar |
| " tti  landline made tr " | ng continues continue continue |

From this model, we can see that many seeds have words that are making sense and headlines as well.

The funiest and best headline within these 15 headlines will be “council continues to search” to me because it’s like a council is searching for something that is not exist and never found it, but they continues to search for it.

**Summary**:

Most of the time in almost every model, headlines have words that are repeating and repeating until forever no matter the model is simple or complex. Weirdly, in most of the seeds, there are some words that make sense which they are supposed to be some debris. The problem may be cause be the dataset is small for the training model or overfitting or underfitting when training the model. Also, because of the time consuming is huge, I didn’t change optimizer and activation for the model training. In further experiments, I will definitely try more of those and also try to make the model more complex if time allows me to do it.