

1.

1)Best result & training process

n_hiddens is: 100

learning_rate is: 0.0015

n_iter is: 150000

Macro accuracy is: 0.5266

Micro accuracy is: 0.5267

5000 3% (0m 9s) 2.1071 gurigah / af Correct

10000 6% (0m 19s) 2.2014 hellesby / fr Incorrect (fi)

15000 10% (0m 30s) 2.3194 bozacioglu / cn Incorrect (za)

20000 13% (0m 42s) 1.7245 pukala / fi Correct

25000 16% (0m 54s) 2.1559 tifirte / fr Incorrect (fi)

30000 20% (1m 5s) 1.1255 freienfelde / fr Incorrect (de)

35000 23% (1m 15s) 2.0266 elysian fields / fr Incorrect (fi)

. . .

100000 66% (3m 47s) 2.0643 finejdet / fr Incorrect (fi)

105000 70% (3m 58s) 1.1026 izbat umm juwayfil / af Incorrect (fi)

110000 73% (4m 9s) 1.5928 lokariya / fi Incorrect (in)

115000 76% (4m 20s) 1.3343 esfidar / fi Correct

120000 80% (4m 31s) 0.9683 am khudera / de Correct

125000 83% (4m 42s) 0.1897 belczace / za Correct

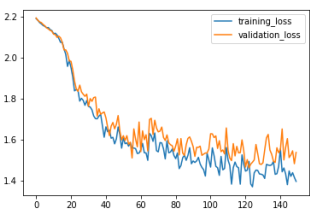
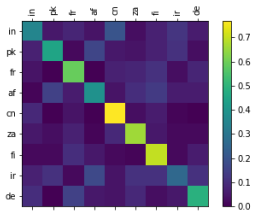
130000 86% (4m 53s) 0.1454 zakaznik / za Correct

135000 90% (5m 4s) 1.4516 tawan / af Incorrect (cn)

140000 93% (5m 16s) 0.7295 kaneriwala / pk Correct

145000 96% (5m 27s) 0.5320 songshuping / cn Correct

150000 100% (5m 38s) 0.6746 einzelberg / de Correct

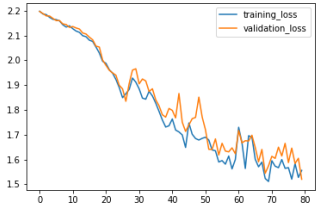
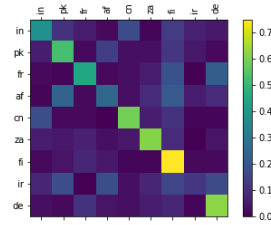
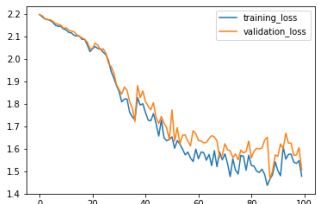
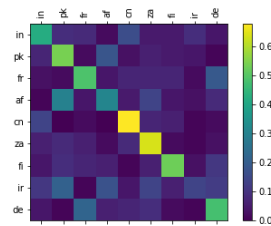
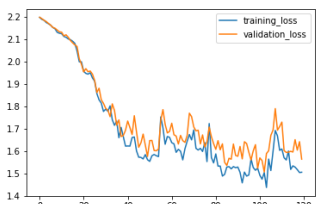
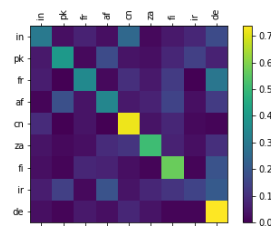
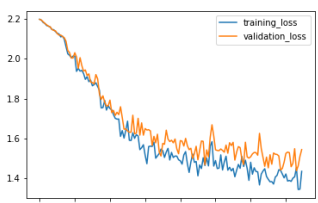
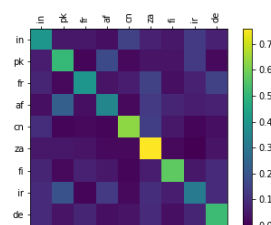
loss	Confusion matrix	Micro_train Macro_train	Micro_val Macro_val
		0.53588 0.5354	0.51777 0.5178

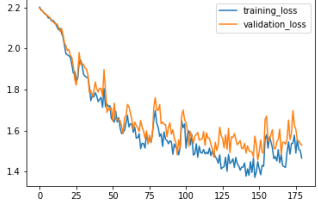
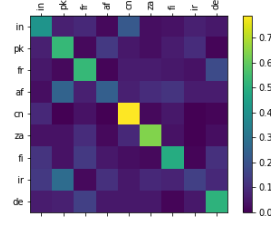
The best result I get is shown above. And some training process and figure also listed above. If the result is correct it will show like "Correct" and if it's not it will

show like "Incorrect" and attached with the correct answer. The confusion matrix shown the Chinese had the best performance under this model (the yellow part) and it also indicate that India city can sometimes be confused with Chinese city. The loss plot proves the model is not overfitting which means the conclusion I get is persuadable.

2) training loop (n_hidden = 80, learning_rate = 0.0015)

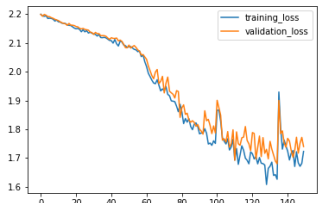
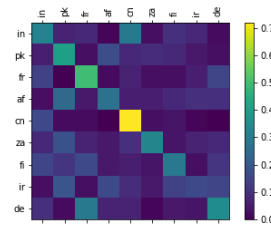
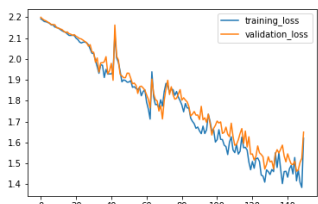
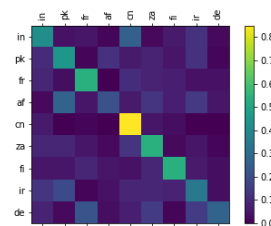
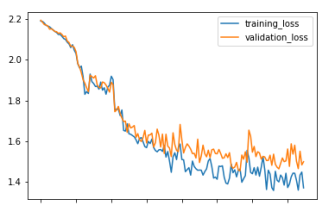
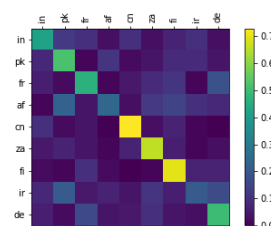
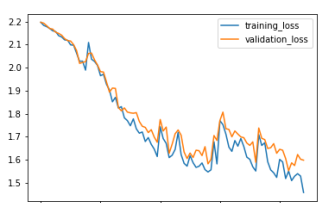
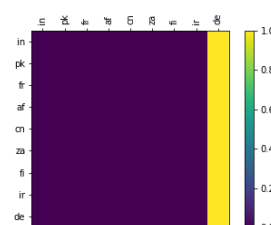
As you can see in the following chart, I choose five different number of the n_iter and get different accuracies. First, 150000 n_iter get the best performance among these numbers and it doesn't mean higher iteration, higher accuracy. For example, it can cause overfitting. (From the picture which means, the different between the two lines is getting bigger and bigger) like "loop = 180000" which have a good performance in training data but poor performance in validation data.

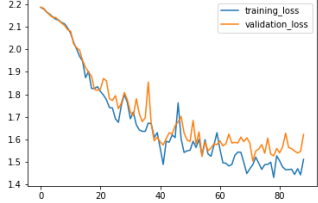
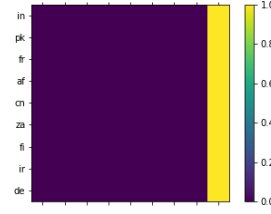
loop	loss	Confusion matrix	Micro_train Macro_train	Micro_val Macro_val
80000			0.49595 0.4959	0.47777 0.4778
100000			0.49294 0.4928	0.47000 0.47
120000			0.48325 0.4836	0.45333 0.4533
150000			0.50784 0.5076	0.49444 0.4944

180000			0.50748 0.5072	0.47666 0.4767
--------	---	---	-------------------	-------------------

3)learning rate(5 different) (n_hidden = 80,n_iter = 150000)

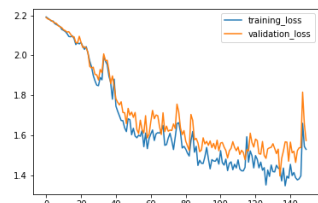
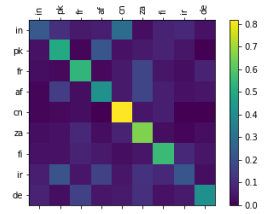
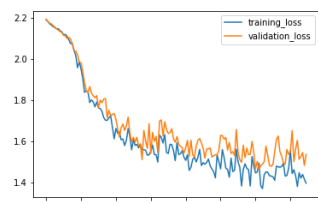
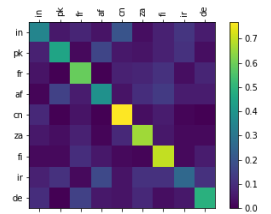
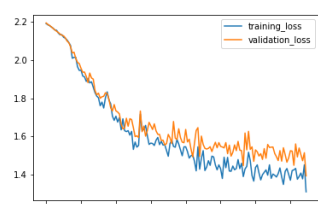
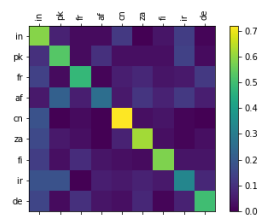
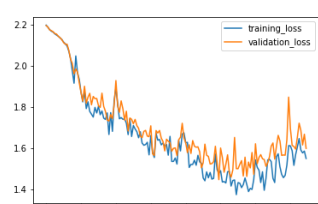
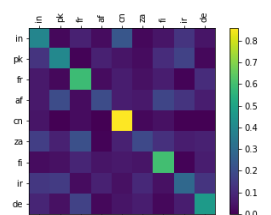
Learning rate is a hyperparameter that controls how much we are adjusting the weights of our network with respect to the loss gradient. If we set it too large it will make the result exceed the optimal value. On the contrary, if we set it too small it will make the falling speed too slow. As we can see in the following chart 0.0015 get the best performance. 0.0025 and 0.002 has bad performance because the NAN result.

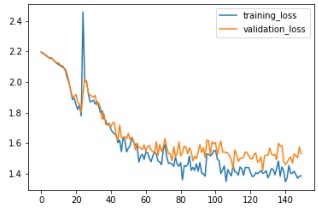
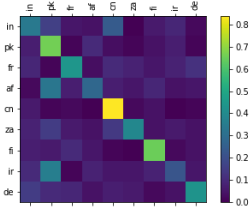
learning rate	loss	Confusion matrix	Micro_train Macro_train	Micro_val Macro_val
0.0005			0.402 0.4028	0.37222 0.3722
0.0010			0.47122 0.4718	0.46222 0.4622
0.0015			0.52477 0.5244	0.49444 0.4944
0.0020			0.111 0.111	0.111 0.111

0.0025			0.111 0.111	0.111 0.111
--------	---	--	----------------	----------------

4)hidden layer(5 different) (learning rate = 0.0015, n_iter = 150000)

In traditional feed-forward neural networks, a hidden layer neuron is a neuron whose output is connected to the inputs of other neurons and is therefore not visible as a network output (hence the term hidden layer). The number of hidden layer we choose really depends on the complexity of the problem we are trying to solve. So it doesn't means more hidden layer lead to higher accuracy. It can also cause overfitting. In the following chart hidden layer = 100 has the best performance on both training data and validation data. And the difference between 100 and 120 seems very small.

hidden layer	loss	Confusion matrix	Micro_train Macro_train	Micro_val Macro_val
80			0.49458 0.4945	0.48444 0.4844
100			0.53588 0.5354	0.51777 0.5178
120			0.53067 0.5309	0.51222 0.5122
140			0.46798 0.4691	0.44333 0.4433

160			0.51218 0.5125	0.47222 0.4722
-----	---	--	-------------------	-------------------

2.

I choose Donald Trump speeches as my dataset for this question. The following examples are generating by my model:

Text generate:

I use "Wh" as the start of the sentences to generate 100 characters sentences during the training.

`n_epochs = 2000`

`hidden_size = 100`

`lr = 0.005`

`temperature=0.8`

the result is not good, it doesn't make sense. But the last one is getting much better than before.

[2m 54s (1800 90%) 1.3399]

Wh I'LL LOcISAIS SONKE AND THIS HEMLIK. WE RERE LPONEN SAY, ON LAR BEVE MTOPPPLE FOLI NGGEW, IS I DON'T

[3m 4s (1900 95%) 1.8671]

Wh evank it to much the tim the sadidnerstose, they're going to never so many a devas tanation big in t

[3m 14s (2000 100%) 1.3930]

WhN'T le going to say, what you know "Or they getions lade - are Iraq STer in I've she gett that.

`n_epochs = 5000`

`hidden_size = 100`

`learning_rate = 0.005`

`temperature=0.8`

I change the `n_epochs` into 5000, but it doesn't help. The final result still like the first one, maybe few words can make sense but the whole sentence is still confused. And the perplexity increased also indicate that I should adjust other parameters.

[7m 51s (4800 96%) 1.6853]

What's trememing he compardest a but I'm coming over the liven. The world the China.

A great a quuent

[8m 1s (4900 98%) 1.5896]

What are back for I could then they just. Give think they people that call. Gold "Ray
sourn. It's just

[8m 11s (5000 100%) 1.7959]

Whiling unwis, then I went ups. Thising timets a dealalse, and in that where Plam - I
'm looks and it a

```
n_epochs = 2000
```

```
hidden_size = 100
```

```
learning_rate = 0.0015
```

```
temperature=0.8
```

This time I lower the learning rate into 0.0015 and it getting better for the
perplexity part and for the generating part more phase making sense now. (like "we
have", "we're going to be" etc)

[2m 51s (1800 90%) 1.6910]

Wh I We'm going to the do thated to that - they fringing to said, we're going to be ar
e in to peron an

[3m 0s (1900 95%) 1.4725]

Why lot trisalle. We're it well there's so me whe spen and the some that people. mou
ld to be deated s

[3m 9s (2000 100%) 1.4654]

Whe ulons. Chishing mear, we have and that's a lot what know very tone the artertula
very to be the fe

```
n_epochs = 2000
```

```
hidden_size = 100
```

```
learning_rate = 0.0015
```

```
temperature=1
```

I increase the temperature because the tutorial says every time a prediction is made
the outputs are divided by the "temperature" argument passed. Using a higher number
makes all actions more equally likely, and thus gives us "more random" outputs. Using
a lower value (less than 1) makes high probabilities contribute more. As we turn the
temperature towards zero we are choosing only the most likely outputs. So I increase
it into 1 and it does getting better than the first one and similar to the third
one.

[2m 52s (1800 90%) 1.5423]

Wh the like this my bogetial, what's agreastecrace hen get evan avance. Thinget and c
omsed tell us. It

[3m 1s (1900 95%) 1.9181]

Wher speny.

And is my stablielby for arour comossing is mother quessts -- therer rich building "I
t.

[3m 11s (2000 100%) 1.6397]

When they see sayouse cominds of I was it it going to ragrial may back hour numped.
These and bush an

Validation:

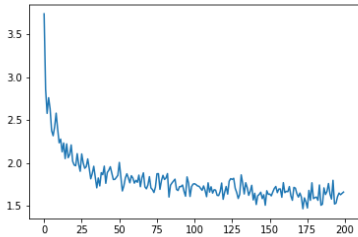
I search on the Internet finding 5 sentences which Donald Trump said in his speech
es but not including in the txt file(1-5) and another 5 sentences from Obama spec
hes which have the some similar word(1-6):

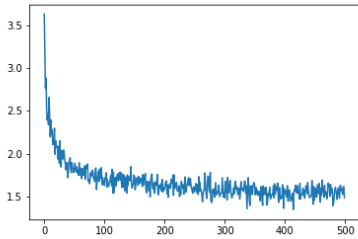
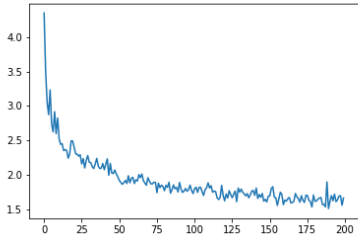
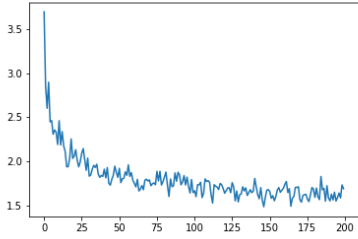
1. In other words, the United States is stronger, safer, and a richer country than it
was when I assumed office less than two years ago.
2. We are standing up for America and for the American people. And we are also standi
ng up for the world.
3. With support from many countries here today, we have engaged with North Korea to r
eplace the specter of conflict with a bold and new push for peace.
4. Ultimately, it is up to the nations of the region to decide what kind of future th
ey want for themselves and their children.
5. Every solution to the humanitarian crisis in Syria must also include a strategy to
address the brutal regime that has fueled and financed it: the corrupt dictatorsh
ip in Iran.
6. But the important point I want to make here is that we already are deeply engaged
in trying to bring about a solution in Syria.
7. For all the maps plastered across our TV screens today, and for all the cynics who
say otherwise, I continue to believe we are simply more than just a collection of
red and blue states. We are the United States.
8. Now, I'm not running for office anymore, so let me just present the facts. I prom
ised that 2014 would be a breakthrough year for America. This morning, we got mor
e evidence to back that up. In December, our businesses created 240,000 new jobs.
I think we should have further broad-based debate among the American people. As

I've said before, I do think that the episode with the unaccompanied children changed a lot of attitudes.

9. And we stand for freedom and hope and the dignity of all human beings. And that's what the city of Paris represents to the world, and that spirit will endure forever -- long after the scourge of terrorism is banished from this world.
10. And we stand for freedom and hope and the dignity of all human beings. And that is what the city of Paris represents to the world, and that spirit will endure forever -- long after the scourge of terrorism is banished from this world.',
11. really like their strong coffee But it's more than that. There's the whole experience First I have to deal with that line up. Depending on the location and time of day Your gonna find about 6 to 8 people in that line up.
12. I never even recognize when I look like someone. I'll be out somewhere and be all like....hahaha! Look at that guy, who does that dork look like?" and then sadly realize... "OH MY GOSH!...that freak looks like me!" eh....what are you gonna do?
13. This skit is also known as "THE GOLF PARADOX" - enjoy:I don't understand golf.The entire point of golf is to play as little golf as possible.To make as few shots as possible.
14. I am going straight to the top you guys. I got my first gig lined up this weekend. Come out if you can. Its the Peachtree 20__ talent showcase. It's gonna be INSANE *pause* in the membrane!My mom, i mean my manager Sheila said i got some real fresh talent!But my career as a rapper really depends on you guys, the fans.
15. None of this would be possible without you guys! and I really need you guys to spread the word.I have a lot of catching up to do since my former partner, marshall s tole my song rap god.Until next time, stay wicked!

And I use these to calculate the perplexity

data	loss	Perplexity(1-5)
<pre>n_epochs = 2000 hidden_size = 100 learning_rate = 0.005 temperature=0.8</pre>		No. 1: 5.95879 No. 2: 3.25370 No. 3: 6.55415 No. 4: 6.52738 No. 5: 10.5454

<pre>n_epochs = 5000 hidden_size = 100 learning_rate = 0.005 temperature=0.8</pre>		No. 1: 6.32819 No. 2: 3.17554 No. 3: 6.41887 No. 4: 6.66983 No. 5: 12.2718
<pre>n_epochs = 2000 hidden_size = 100 learning_rate = 0.0015 temperature=0.8</pre>		No. 1: 6.34050 No. 2: 3.33506 No. 3: 6.66575 No. 4: 6.18419 No. 5: 9.79403
<pre>n_epochs = 2000 hidden_size = 100 learning_rate = 0.005 temperature=1</pre>		No. 1: 5.28735 No. 2: 3.51779 No. 3: 7.11666 No. 4: 5.39794 No. 5: 9.71912

And then I choose the best performance (n_epochs = 2000, hidden_size = 100, learning_rate = 0.00, temperature = 1) to test on validation(6-15). 6-10 sentences come from Obama speeches which is like Donald Trump because they are both presidents and I choose the same topic and short sentences. 11-15 come from a comedy script which is dissimilar to a public speech. I got the following result: As you can see in the following chart, the speeches have the similar perplexity but the Comedy script have the big different from the first two.

Donald Trump	Obama	Comedy script
No. 1: 5.5935	No. 6: 5.0832	No. 11: 5.9222
No. 2: 3.7860	No. 7: 6.7554	No. 12: 10.3402
No. 3: 6.6715	No. 8: 6.7374	No. 13: 13.6081
No. 4: 5.9655	No. 9: 6.3943	No. 14: 11.2428
No. 5: 8.8633	No. 10: 6.0965	No. 15: 8.2807