6006 Dissertation (2024-2025 Semester A):

An Attempt on Transformer Realization

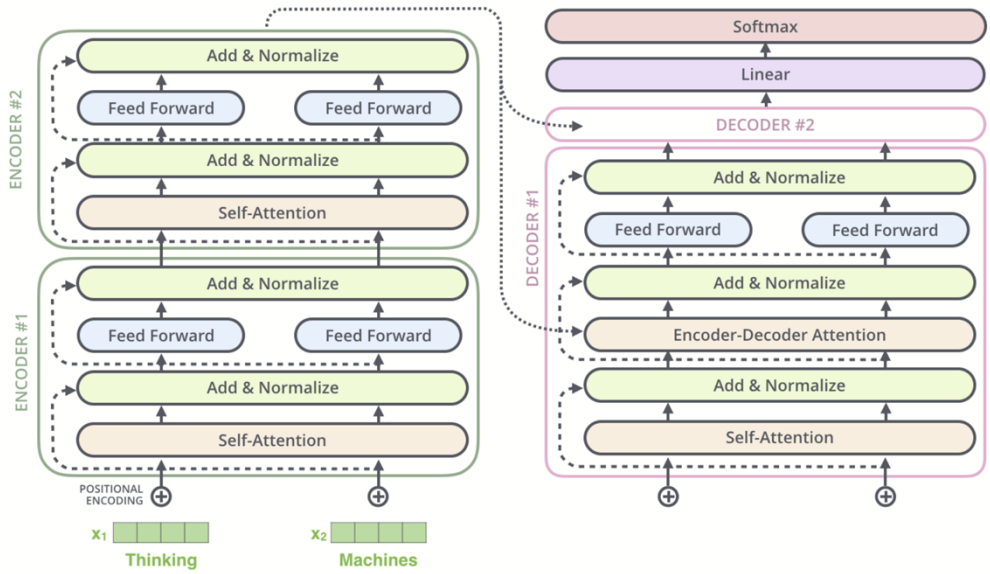
Name: WANG Haoyu ID: 59066216 Email: hwang845-c@my.cityu.edu.hk

Part I. Project Destination

Before midterm of semester A, we talked about the principles of Transformer. To summary, it is a huge natural language processing model applied all kinds of mathematic skills. By using the core technique called “Attention Mechanism”, which means to calculate the relations between one token and the others, it predicts the translated outcome of a sentence. The main goal of this project is to learn the mathematic mechanism of Transformer, try to realize the Transformer described in the paper, explore the properties of Transformer’s output, build further applications based on Transformer.

Part II. A Brief Introduction to Transformer

The illustration below is the sketch of Transformer, most of the layers included in it is shown in the picture.



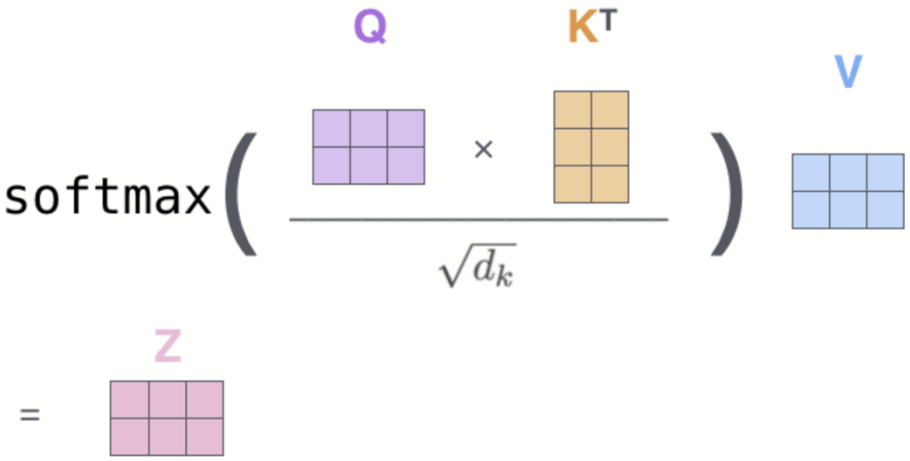
After tokenization, each token is embedded into a vector, add positional encoding, and treated as the input of the encoder.

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{2i}/d_{model}}\right)$$

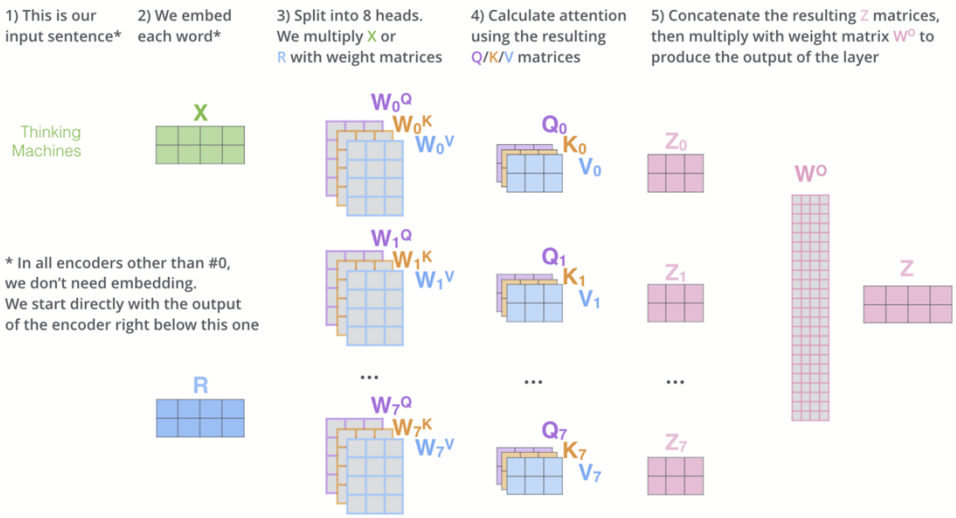
$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{2i}/d_{model}}\right)$$

Then we calculate the Q, K, V matrix using weights

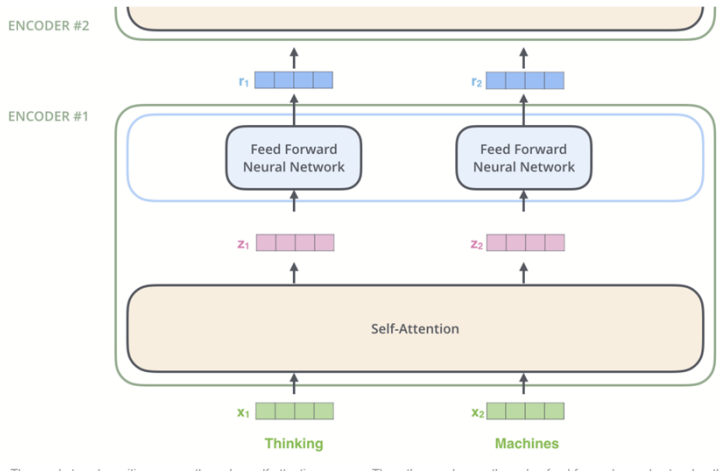
Then calculate the output of the self-attention layer.



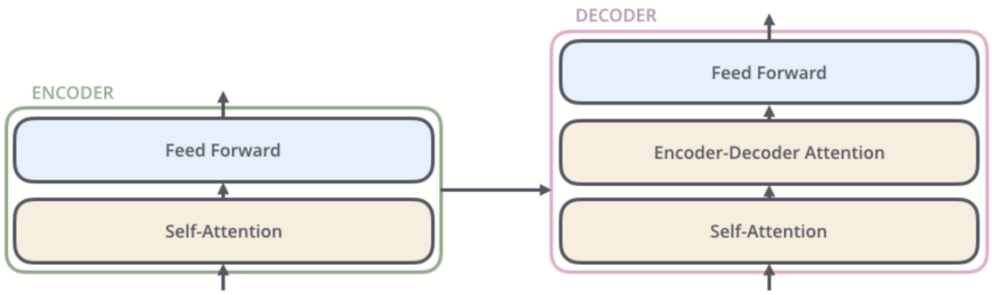
To get better result, we apply multiple heads, so the process may seem like this:



After feed the outputs to the paralleled Feed-forward Layer, we build an encoder block, the encoder blocks altogether form the encoder of the Transformer.



Similar to encoder, a decoder block consists of a self-attention layer, an encoder-decoder layer and a feed-forward layer.

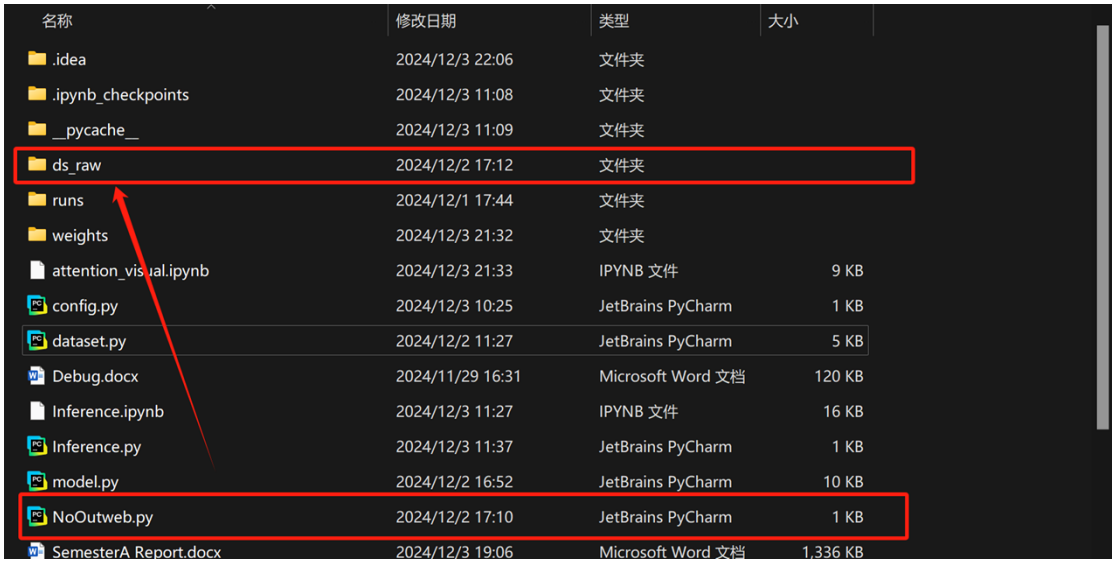


Part III. Research Progress

I have built a Transformer of translation so far, it aims at translating English into Italian. I have trained the model for 50 epochs and the result is pretty good, it can almost translate the sentence correctly. A flaw of the present model is there are some padding tokens appear in the outcome but it doesn't affect understanding.

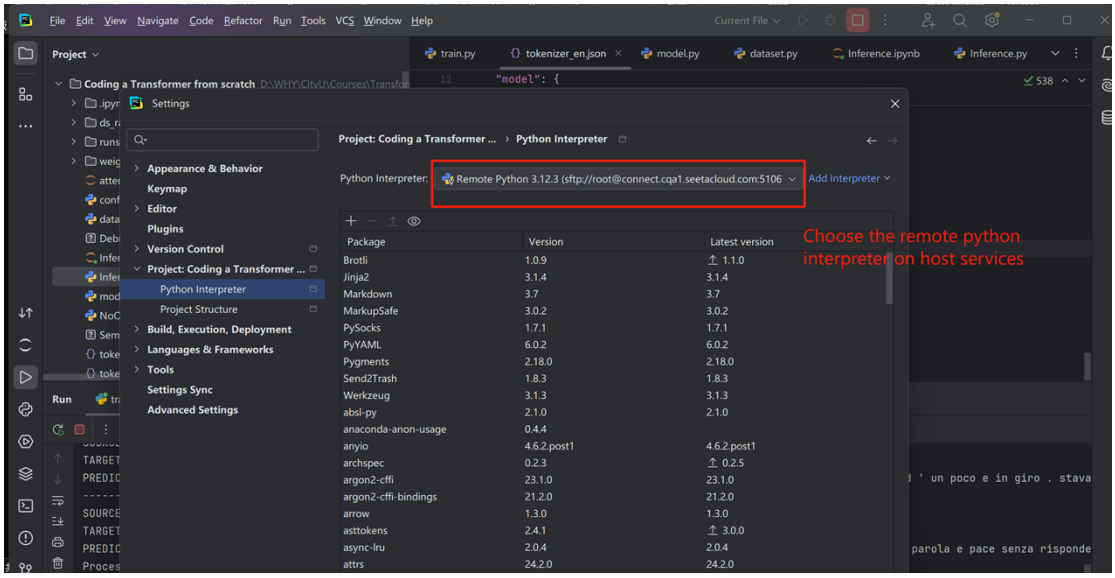
Part IV. Run the Code and Cautions

1. First of all, because we use the dataset on Hugging face, which is a foreign website, and the host server in China, which cannot access the foreign website, we need to download the dataset to local first, and then upload the dataset to our host server. For this step, run NoOutWeb.py to download the dataset.

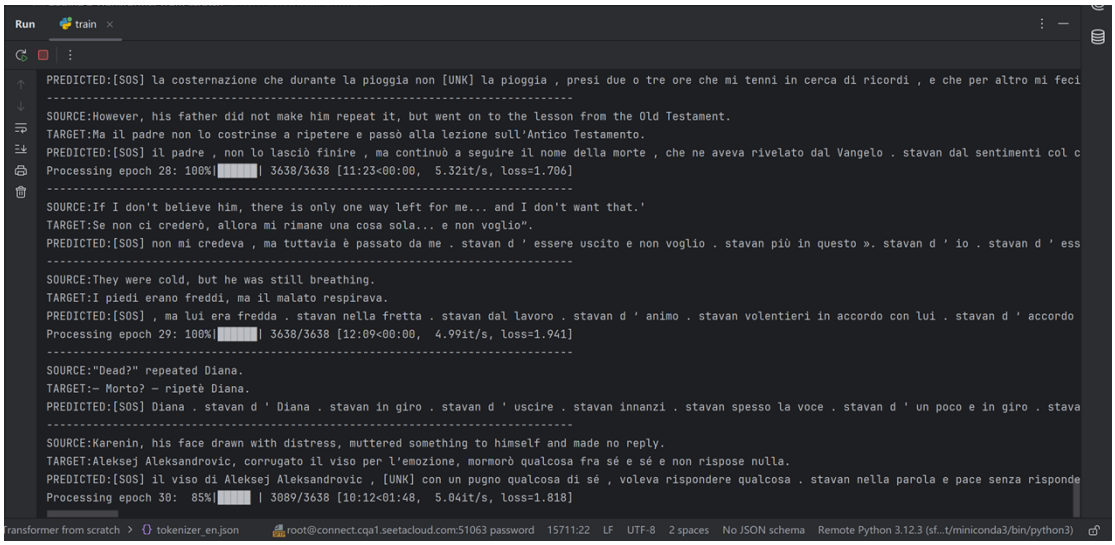


2. When we first run the main code (train.py), you may find that the dictionaries of the tokenizers of both languages miss tokens: [SOS], [EOS], [PAD], [UNK]. But don't worry, I didn't find corresponding function in python to add these tokens to the dictionaries by code, so I manually add these tokens to the dictionaries of tokenizers in json file. You can use the tokenizers provided by me (tokenizer_en.json, tokenizer_it.json), or you can generate them by yourself and add these missing tokens to the dictionaries.

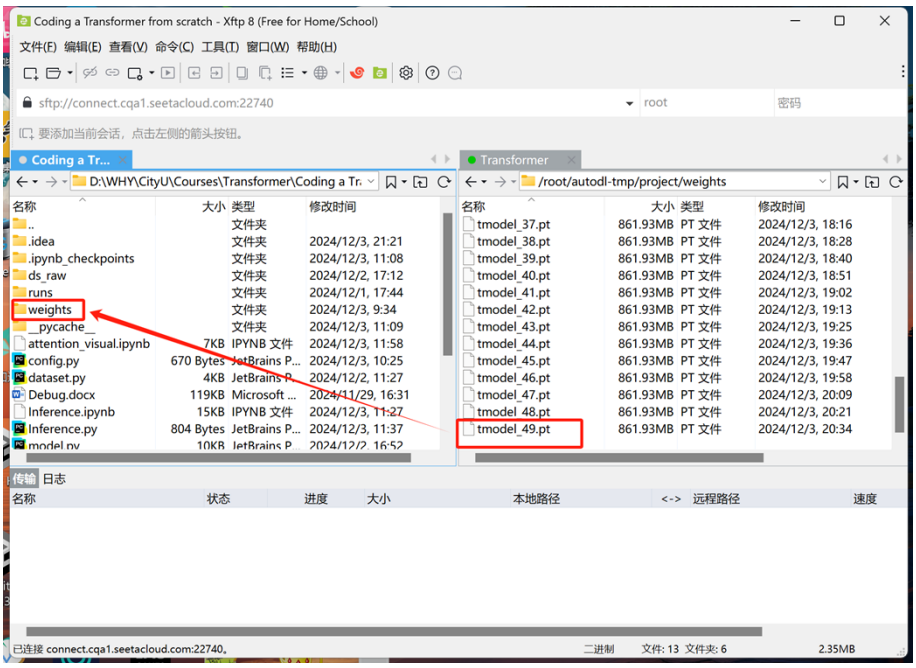
3. Choose the Python Interpreter. There may be other ways to train the model, but I used 'cuda' device.



4. Then run the “train.py”. You can see the progress of the training, there are also test translation at the end of the epoch.



5. After finishing the training progress, download the weights from host server. (Here we use Xftp8 to realize file transfer)

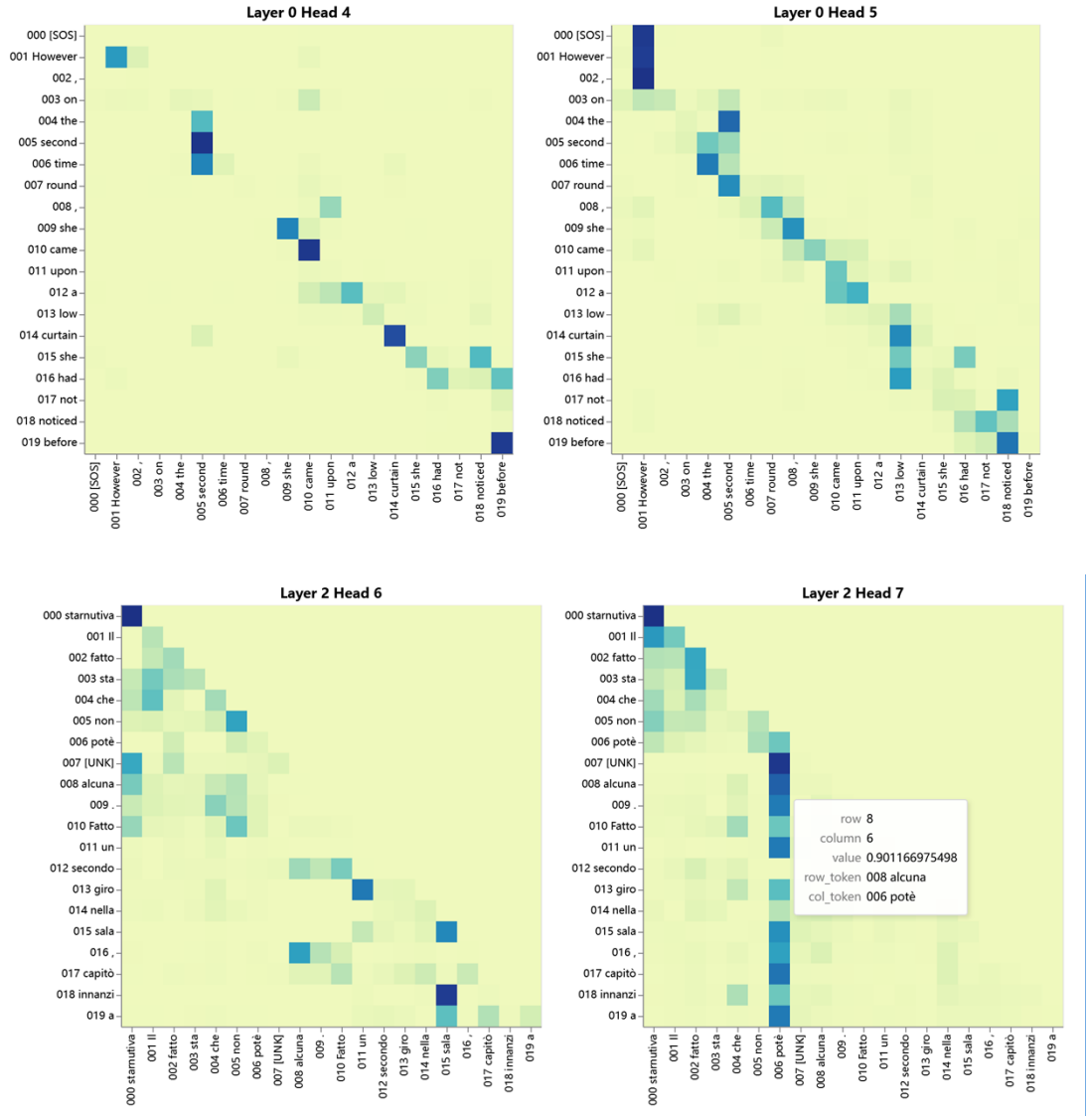


6. Then run the validation via jupyter notbook. Gain output like below:

```
--------------------------------------------------------------------------------
SOURCE:George got quite huffy; but, as Harris said, if he didn't want his opinion, why did he ask for it?
TARGET:Giorgio si stizzì; ma, come Harris gli disse, se non voleva la sua opinione, perchè gliela domandava?
PREDICTED:[SOS] : Giorgio , ma Harris disse , come se avesse potuto [UNK] la sua opinione , perchè gliela domandava ? stavan da
l fratello . stavan da questo ? stavan di sapere ? stavan su la sua fede . stavan da lui in modo . stavan ?". stavan da lui . s
tavan . stavan . stavan e disse : stavan . stavan . stavan . stavan . stavan . stavan . stavan . stavan dal suo sguardo . stava
n . stavan . stavan . stavan . stavan . stavan . stavan . stavan . stavan . stavan . stavan . stavan . stavan dal significato c

--------------------------------------------------------------------------------
SOURCE:'When I'm a Duchess,' she said to herself, (not in a very hopeful tone though), 'I won't have any pepper in my kitchen a
t all.
TARGET:“Quando sarò Duchessa, — si disse (ma senza soverchia speranza), — non vorrò avere neppure un granello di pepe in cucina
.
PREDICTED:[SOS] : “ La Duchessa ” si disse ( ma senza mia pena ), non vorrò avere neppure un granello di pepe in cucina . stava
n d ' un camino che io non ho mai avere in cucina . stavan affatto . stavan d ' un camino . stavan innanzi . stavan d ' intrapr
endere aver fretta . stavan . stavan d ' un colpo , e poi farò vedere a me . stavan . stavan d ' [UNK] . stavan . stavan di vis
ta . stavan . stavan . stavan . stavan . stavan . stavan . stavan . stavan . stavan . stavan . stavan . stavan . stavan . stava
n . stavan . stavan . stavan . stavan . stavan innanzi . stavan innanzi . stavan . stavan . stavan . stavan . stavan . stavan i

--------------------------------------------------------------------------------
SOURCE:'Yes, I will go,' said Anna, recovering and rousing herself; 'and if a telegram comes during my absence, send it to Dary
a Alexandrovna's...
TARGET:— Sì, andrò — disse Anna tornando in sé e alzandosi. — E se verrà un telegramma quando non ci sarò, mandatemelo da Dar’j
a Aleksandrovna.
PREDICTED:[SOS] — disse Anna , alzandosi — e [UNK] in modo del mio .... — E se verrà un telegramma quando ... stavan nella mia
moglie ... stavan tardi . stavan tardi . stavan tardi . stavan tardi . stavan dai Karenin . stavan da Dar ’ ja Aleksandrovna .
stavan d ’ andare a parlare . stavan di parlare . stavan . stavan .... stavan . stavan . stavan dal mio vettura . stavan . stav
an ! stavan d ’ aiuto . stavan d ’ intraprendere ciò .... stavan . stavan . stavan . stavan . stavan . stavan d ’ un biglietto

--------------------------------------------------------------------------------
SOURCE:But I was no apostle,--I could not behold the herald,--I could not receive his call.
TARGET:— Ma non ero apostolo, non potevo scorgere il messaggero e non potevo ricevere il suo ordine.
PREDICTED:[SOS] non era vero , non potevo fare attenzione al libro , non potevo ricevere nessuna cugina . stavan da lui . stava
n dal agio . stavan dal viso . stavan di chiamare le prime le prime parole . stavan dal viso . stavan e mi pregò di occuparsi .
stavan dal viso . stavan e non posso esser testimone di non avevo ancora il ricevere il volto . stavan dei passi . stavan . sta
van dal volto . stavan dei passi in faccia a ricevere sotto il viso . stavan del apparivano vivo . stavan . stavan . stavan d '


--------------------------------------------------------------------------------
SOURCE:One afternoon (I had then been three weeks at Lowood), as I was sitting with a slate in my hand, puzzling over a sum in
long division, my eyes, raised in abstraction to the window, caught sight of a figure just passing: I recognised almost instinc
tively that gaunt outline; and when, two minutes after, all the school, teachers included, rose _en masse_, it was not necessar
y for me to look up in order to ascertain whose entrance they thus greeted.
TARGET:Un pomeriggio, mentre ero seduta con la lavagna sulle ginocchia e mi arrabattavo per fare un'addizione lunga, alzai gli
occhi per guardare verso la finestra e vidi passare una figura, che riconobbi istintivamente. Due minuti dopo, tutta la scuola
si alzava in massa e non ebbi bisogno di guardare per capire chi era salutato a quel modo.
PREDICTED:[SOS] a Lowood feci che avevo appena le poche bicchiere di Lowood con la lavagna che mi [UNK] con fare uno sforzo per
guardare la finestra ; vidi che una finestra si era sollevata e vidi una faccia [UNK] , che nessuno mi aveva fatto passare la s
cuola , quando tutte le persone si era sala a guardarla per capire tutto ciò che era venuto in sala . stavan di far cessare la
scuola . stavan spesso in salotto , quando udii andare a visitare tutte le persone . stavan lontani . stavan lontani . stavan l
ontani . stavan lontani dalla sala . stavan di vedere chi sa dove . stavan . stavan . stavan . stavan . stavan innanzi
. stavan in salotto . stavan a dare in salotto . stavan lontani . stavan . stavan . stavan . stavan a parlare . stavan in salot
```

```
--------------------------------------------------------------------------------
SOURCE:A candle just brought in gradually lit up the study and its familiar details became visible: the stag's horns, the book
shelves, the looking-glass, the hot-air aperture of the stove with its brass lid, which had long been in need of repair, his fa
ther's couch, the large table on which were an open volume, a broken ash-tray, and an exercise-book in his handwriting.
TARGET:Lo studio fu illuminato a poco a poco da una candela che vi portarono. Cominciarono a comparire i noti particolari; le c
orna di cervo, gli scaffali coi libri, lo specchio, la stufa con la bocca di calore che da tempo doveva essere riaccomodata, il
divano del padre, il grande scrittoio, sullo scrittoio un libro aperto, un portacenere rotto, un quaderno con la propria scritt
ura.
PREDICTED:[SOS] una candela in dispensa a poco a poco il studio e alla figura dello studio dei contadini , che si avvicinavano
a quei libri , l ’ un libro di velluto , guardato le estremità del portacenere [UNK] , con la tavola grande , coll del bisogno
del portacenere allegro , ancora più [UNK] , un libro di seno che aveva il libro aperto , un libro aperto , un libro aperto , u
n libro sulle sue braccia . stavan di sangue , aveva [UNK] una lettera . stavan di lato e un biglietto in modo da una lettera .
stavan di pena . stavan di parlare con la propria scrittura . stavan . stavan . stavan nella sua azione . stavan nella sua fret
ta e un biglietto . stavan da una lettera . stavan nella sua fretta . stavan da una lettera . stavan d ’ [UNK] . stavan d ' [UN
```

7. You can also visualize the attention heat map by run the Attention Visualization.



#Above is the application of my Transformer. There may be many problems generated while running the codes because of different environments and

settings, so if you have any problems about the application, you can contact me via email.

Part V Future Prospect

A Transformer on translation between English and Italian has been built, but there are still some flaws of the model. After correcting these flaws, I want to talked about the properties of Transformer. For example, there are multiple heads of Transformer, “Is function of each head meaningful and credible?”, “Could the outputs of a token through all heads be the same?”.

After the exploration of these questions, I want to apply Transformer in some other scenes. For example, extend it to translations between other languages, and even to other categories such as sentiment analysis, semantic analysis, image identification and so on. There is no doubt that Transformer has that kind of potential to solve problems in lots of categories, but this requires to learn more about the Transformer and build models far more complicated. We can build the Transformer into a super powerful monster, but before that, there are still long roads for us to walk down.