Artificial Intelligence Final Report Assignment 問題1 (Problem 1)

レポート解答用紙 (Report Answer Sheet)

Group Leader

学生証番号 (Student ID): 18520952

名前(Name): Ngo Le Hieu Kien

Group Members

学生証番号 (Student ID): 18521064

名前(Name): Ngo Dinh Luan

学生証番号 (Student ID): 18521029

名前(Name): Le Hoang Long

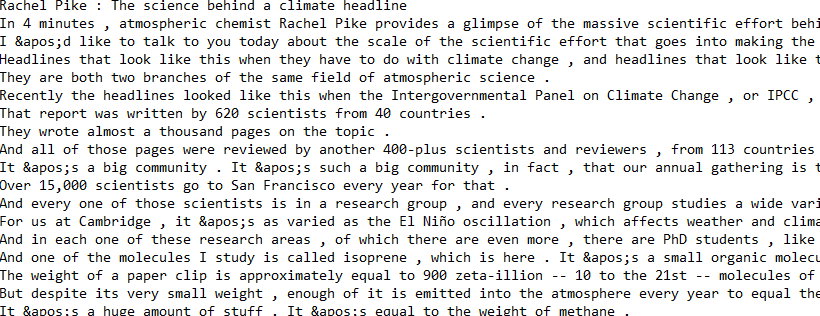
学生証番号 (Student ID): 18521024

名前(Name): Le Hoang Long

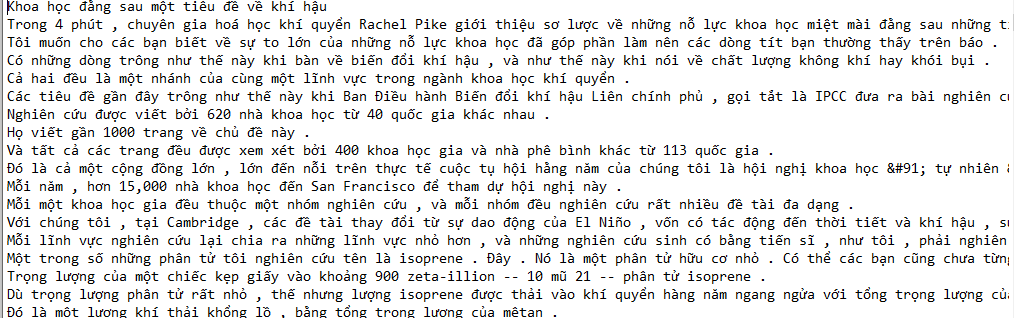
問題1 (Problem 1)のレポート

1. Introduce datasets

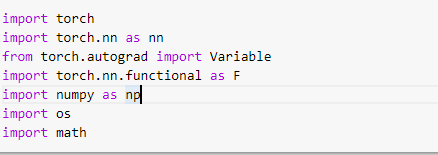
+ Train\_en sets :



+ train\_vn sets :

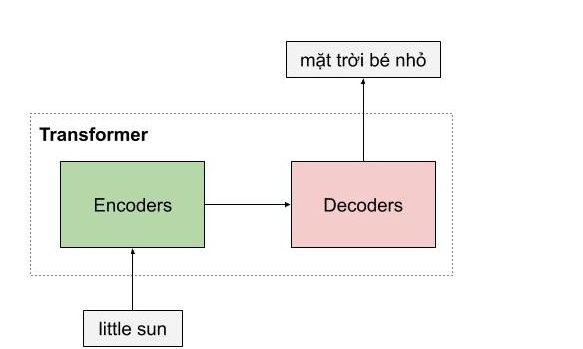


Libraries need to be prepared :

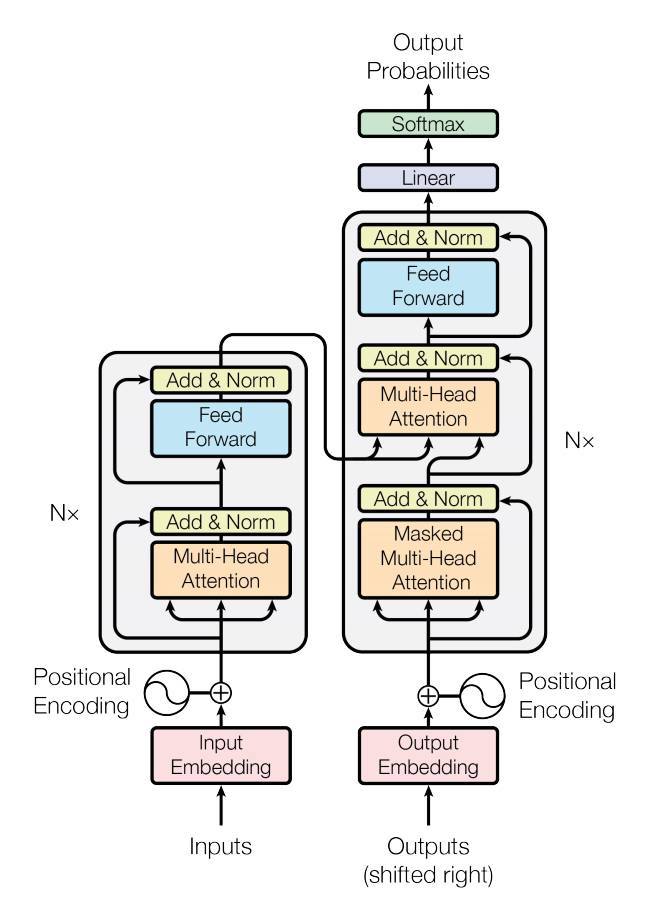


1. Transformer
2. Introduce :

Transformer models popularity is undeniable, because it is the base, the core to develop all other famous model like BERT (Bidirectional Encoder Representations from Transformers) - A model use to learn how to present the best vocabulary and it had been a big step for the NLP community in the year of 2019. Google is also use BERT model for their searching engine. To start understand more about BERT, you have to know about the Transformer model.

Model overview :

1. Detailed model .

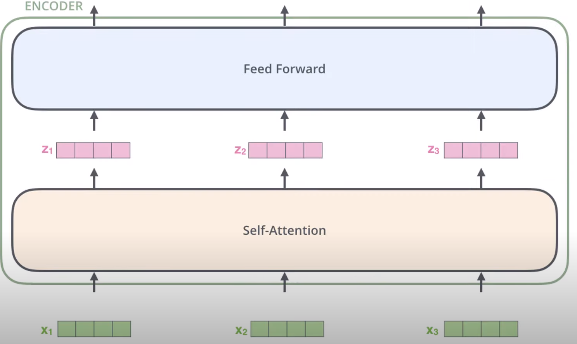


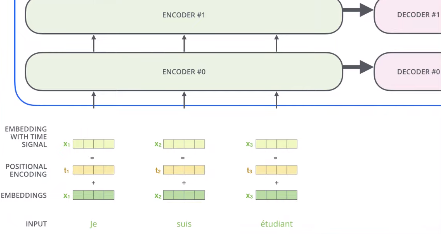
2.1 ENCODER

Transformer models encoder can contain a lot of similar encoder layer, each encoder is created by 2 elements which is multi-head attention and feedforward network, as well as skip connection and normalization layer. of these 2 element,you will be more interested in multi-head attention because it is a new layer which is being introduce in this report, and its also help making the different between Transformers model and LTSM model.

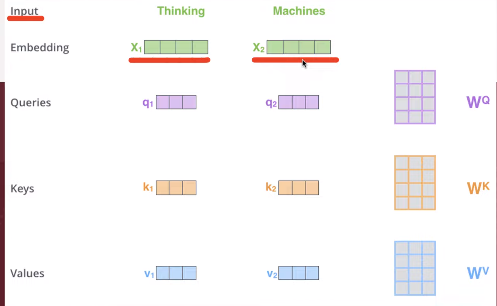
+ Consists of 2 main blocks:

Self-Attention and Feedforward Neural Network (the most important class is Self-Attention)



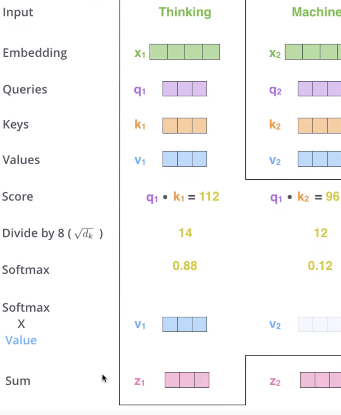
+ Before the input enters the Self-Attention class, the embedding will go through the Positional Encoding class to ensure word order in the sentence

+ Self-Attention class



The query q1 will find the relationship with the remaining words. The W will be learned during the relationship training with words.

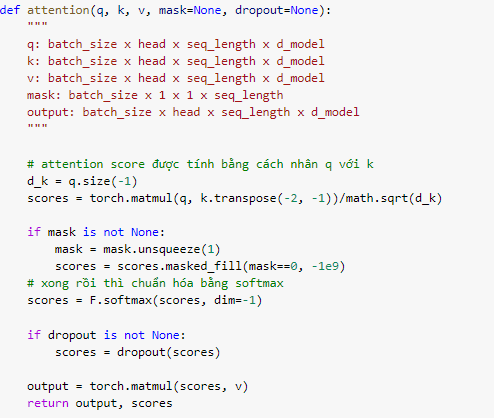
Calculation :

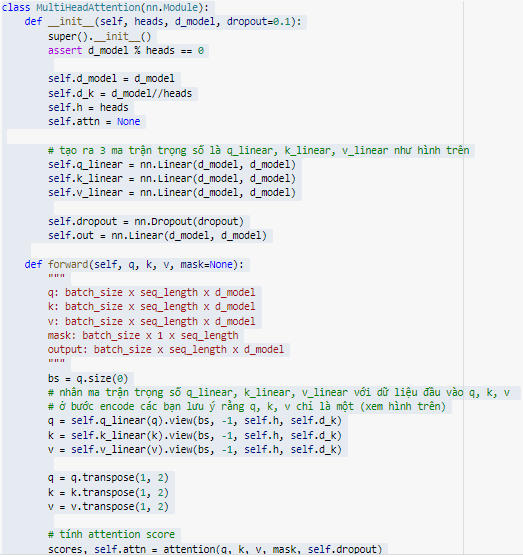


Vector V1 in bold color --🡪 q1 has a large relationship with k1

Vector V2 in light color--🡪q1 has a low relationship with k2

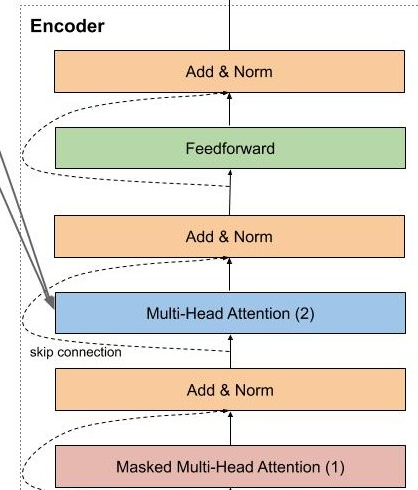
+ Construction :





2.2 DECODER

Decoder performs the function of decoding the source sentence vector into the target sentence, so the decoder will receive information from the encoder as 2 vector key and value. The decoder's architecture is very similar to that of the encoder, except that there is an extra multi head attention in the middle used to learn the relationship between the word being translated and the words in the source sentence.

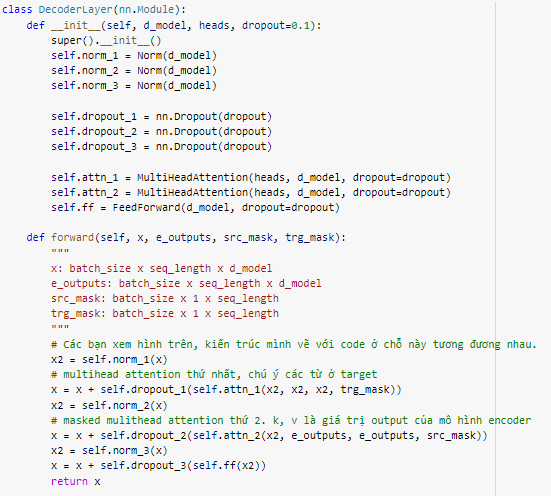
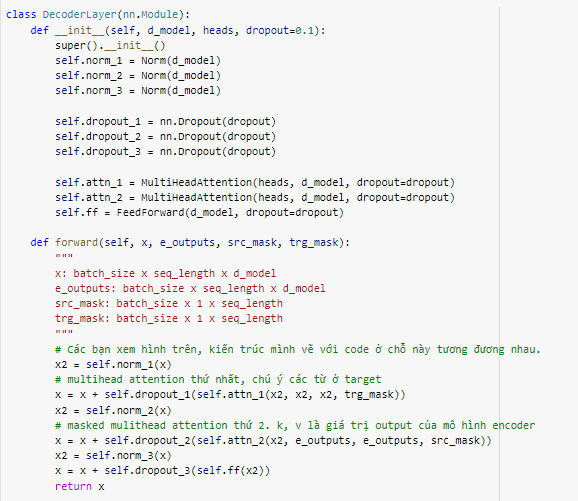


## +Masked Multi Head Attention Layer:

Masked Multi Head Attention is of course the multi head attention that we talked about above, which has the function to encode target sentence words during translation, however, when installing we need to note that we have to hide future words are not yet translated by the model, to do this we simply multiply by a vector containing the values 0.1.

In the decoder there is also another multi head attention function that notices the words in the encoder model, this layer receives the key and value vectors from the encoder model, and outputs from the layer below. Simply because we want to compare the correlation between the word being translated from source.

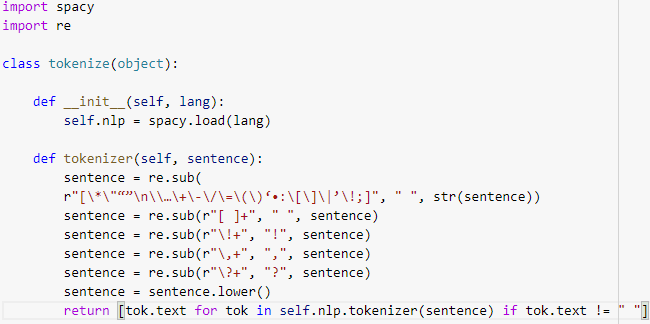
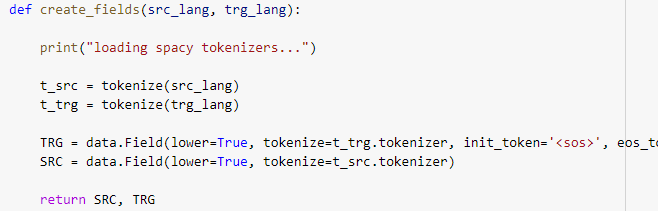
+ Construction :



2.3 Install transformer

## 

1. Data processing





1. Evaluate the model based on the BLEU measure.

BLEU stands for Bilingual Evaluation Understudy, is a method of evaluating a translation based on reference translations. The prerequisite to be able to use BLEU is that you must have one (or more) sample sentences. For the machine translation problem, the sample sentence is the output sentence of the pair of sentences in the data set. BLEU evaluates a sentence by matching it with sample sentences and gives it a scale from 0 (absolute deviation) to 1 (absolute match).

BLEU is known as a method that is simple, easy to understand, low computational cost and similar to human assessment. However, the human factor in sample sentence generation makes BLEU not completely objective.

+The calculation of BLEU is to count the number of n-gram matches between the sample sentence (R) and the evaluated sentence (C) and then divide it by the number of tokens of C. The choice of n depends on the language, task, and specific goal. body. The simplest we can use unigram is n-gram containing 1 token (n=1). Visually, the larger n, the smoother the sentence.

IMG_256

The maximum number of matches is limited by the maximum number of occurrences of n-grams in the sample sentences.

