Rapport pour le projet de
Deep Learning, NLP
Version du 20 Juillet 2022

Introduction to Natural Language Processing
Catch-up 2

Sous la direction de

Marc VON-WYL

Christopher DIAMANA LUTETE
EPITA SCIA promotion 2022

Table des matières

1 Theoritical questions

1.1 Question 1

What is the purpose of subword tokenization used by transformer models ?

❖ ***Part of the answer is in the first part of the course***

The purpose of subword tokenization is to cut words into morphemes, the smallest unit of text containing meaning.

❖ ***What is the effect on the vocabulary size ?***

The subword tokenization limits the size of a vocabulary.

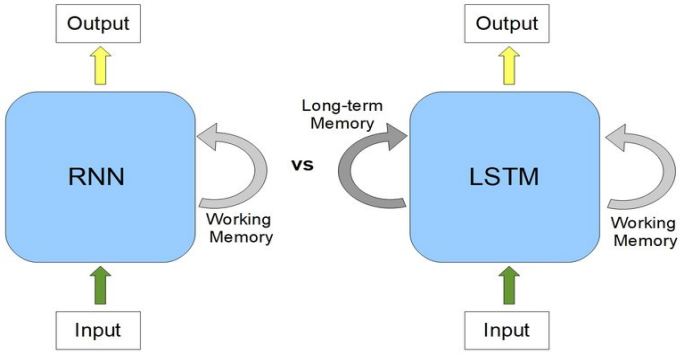
❖ ***How does it impact out-of-vocabulary words (words which are not in the training data, but appear in the test data, or production environment) ?***

The subword tokenization reduce even avoid out-of-vocabulary words because it can create a representation of the unknown words.

1.2 Question 2

What are the differences between an RNN and an LSTM ?

The Recurrent Neural Network (RNN) use their internal state to process sequences of inputs. Long Short-Term Memory (LSTM) is a varient of RNN, with addtional long term memory to remember past data. So the main difference between RNN and LSTM will be how which one maintain information in the memory for the long period of time.



❖ ***What problem is an LSTM trying to solve compared to a basic RNN ?***

LSTM is trying to solve the Vanishing Gradient problem.

1.3 Question 3

When building an encoder-decoder model using an RNN, what is the purpose of adding attention ?

Adding attention when building an encoder-decoder allow the model to focus on certain parts of the input sequence when predicting a certain part of the output sequence.

❖ ***What problem are we trying to solve ?***

Attention mechanism is trying to solve bottleneck problem.

❖ ***How does attention solve the problem ?***

Attention solve the problem by using hidden states at all time-steps of input sequence for better modelling of long-distance relationships.Thus, the decoder can access all the hidden states and not only the final state.

1.4 Question 4

In a transformer model what is the multihead attention used for ?

❖ ***What the purpose of self-attention ?***

The self-attention encodes each word as a function of the other words in the input. Thus, the self-attention will help the network learn to associate words.

❖ ***Why do we use muliple head instead of one ?***

Because rather than learning a single type of relationship between words, multi-head attention proposes to have several heads per layer of self-attention with the expectation that each head will focus on one type of relationship.

1.5 Question 5

In a transformer, what is the purpose of positional embedding ?

Positional embedding allows the model to have information about the order of a sequence. The goal is to give a clue about the position.

❖ ***What would be the problem if we didn't use it ?***

The transformer would have no information regarding the sense of position/order for each word in a sentence.

1.6 Question 6

What is the purpose of having benchmarks to evalute models ?

It provides a way to measure/monitor the progress in the models. But it is also a way to know what to use.

1.7 Question 7

In the BERT model, describe the two tasks used for pre-training (unsupervised) with a few sentences.

The two tasks used for pre-training are the Masked Language Modeling (MLM) and the Next Sentence Prediction (NSP).

- MLM : In this task some the tokens from each sequence are randomly masked by the token [MASK] and the model is trained to predict these tokens using all the other tokens of the sequence.
- NSP : In this task there are two input sequences separated by the token [SEP] token and the model is trained to predict if the second sentence succeeds the first sentence in the corpus. It is a binary classification task.

Are they really unsupervised ?

...

1.8 Question 8

In a few sentences, explain how the triplet loss is used to train a bi-encoder model for semantic similarity ?

- ***The simplest version of the triplet loss.***

...

1.9 Question 9

What is the purpose of using an Approximate Nearest Neighbour method to speed up search ?

The purpose of using an Approximate Nearest Neighbour is to find an estimate of the nearest results.

- ***What does it reduce ?***

...