

MVA / ALTEGRAD 2022 / LAB SESSION 3: NLP FRAMEWORKS

IMANE ELBACHA

22 November, 2022

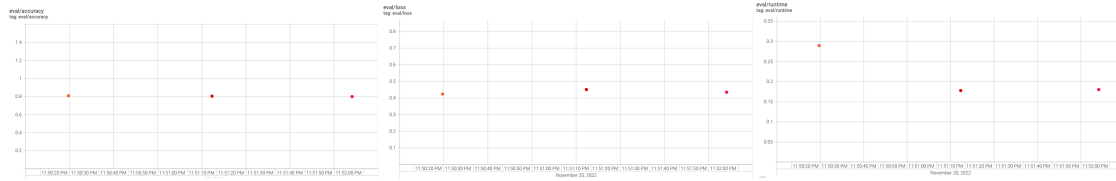
1 Question 1:

Compute the number of parameters of the model manually.

- First layer for the sentence encoder embed tokens: $n_{encoder1} = n_{input} * n_{output} = 32000 * 512 = 16384000$
- Second layer sentence encoder embed positions: $n_{encoder2} = n_{input} * n_{output} = 258 * 512 = 132096$
- self attention layers, there are 4 self attention layers each one with 6 projection: $n_{att} = n_{layers} * n_{proj} * n_{input} * n_{output} = 4 * 6 * 512 * 512 = 6291456$
- In total there are 22807552 trainable parameters

2 Question 2:

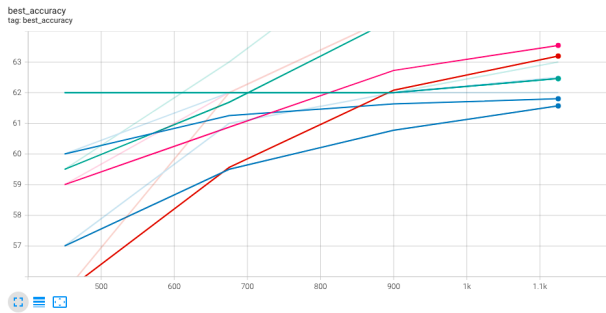
Open Question: make a comparison between the two frameworks. Which one do you prefer?



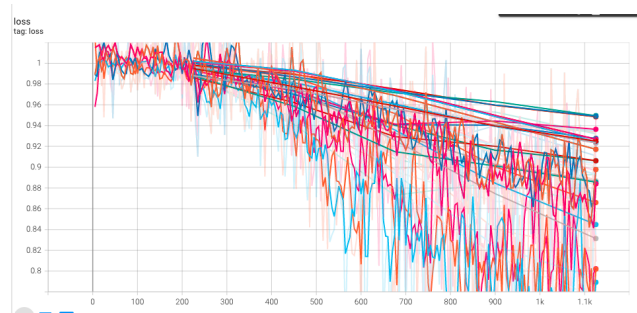
(a) Accuracies of HF models

(b) Loss of HF models

(c) runtime of HF models



(a) Accuracies of Fairseq models



(b) loss of Fairseq models

Commentary:

- Regarding the distribution of the accuracies it is clear that the HF transformers give better and more stable results (but it is important to note that this could be due to differences in model set up)
- the computational time is significantly shorter for HF models as the Fairseq models take longer computational time for the same number of epochs.

- The preprocessing is very simple for the HF transformers (requiring just the conversion of the data into json files) whereas Fairseq transformers require the tokenizing of the input
- Finally, since the implementation of the HF transformers is simple and fast i prefer this Framework