

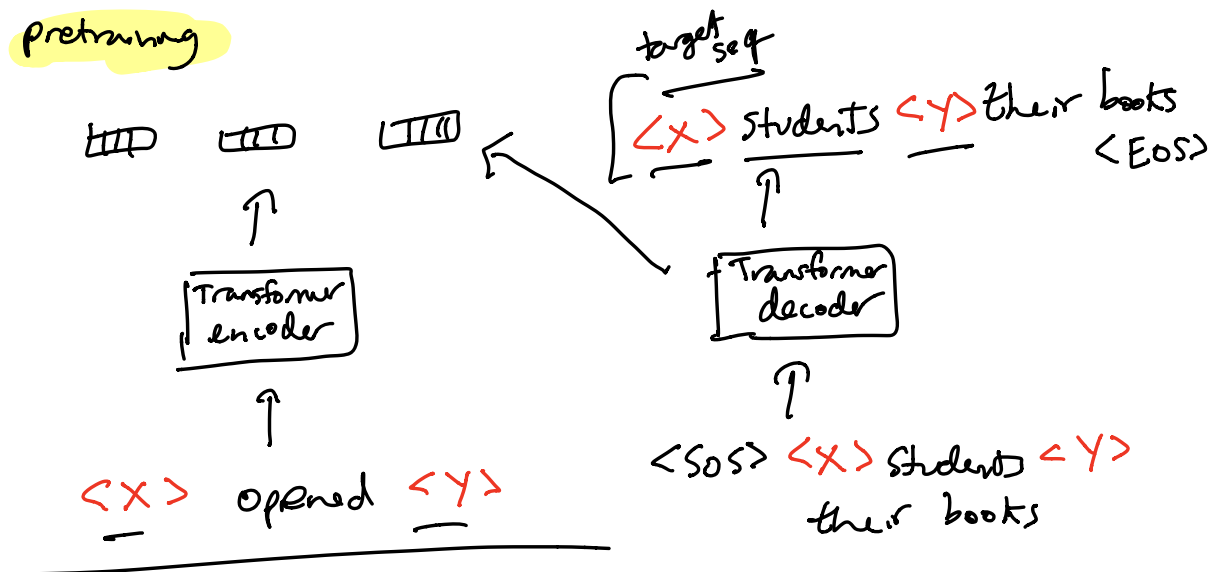
TS: pretrained seq2seq model

- unlike BERT, TS generates text
- encoder/decoder
- span corruption pretraining objective

ex: "students opened their books"

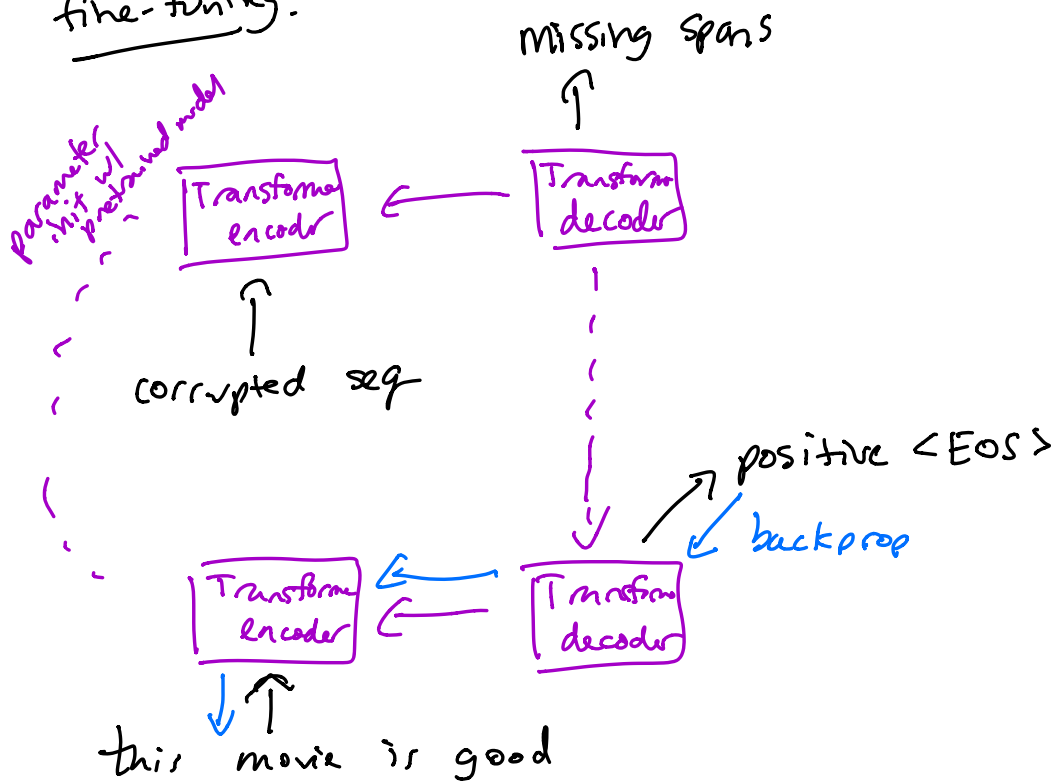
\uparrow Sentinel tokens \uparrow
<X> opened <Y>

pretraining



target is shifted over one token from input

fine-tuning:



- the decoder is generating the text of the label. i.e. it needs to choose "positive" from the entire vocab.
- this is in contrast to BERT, in which you only produce {positive, negative} with a binary classifier