

CS224 - Assignment 4

O.L.

September 2019

1

1.g

The effect: putting $-\infty$ in the padding locations make the softmax value to be zero.

Necessity : the attention mechanism wouldn't accidentally focus on the padded values.

1.i

Corpus BLEU: 22.691071290282704

1.j

In the hierarchical order:

dot product attention \rightarrow multiplicative attention \rightarrow additive attention.

More complex models and therefore takes more resources and time to compute, at the same time have greater capabilities.