gradients/bert/encoder/layer_3

output
18.01%

attention
62.80%

intermediate
19.19%