gradients/bert/encoder/layer_3

output
18.00%

attention
62.86%

intermediate
19.14%