gradients/bert/encoder/layer_0

output
17.92%

attention
63.27%

intermediate
18.81%