gradients/bert/encoder/layer_8

output
21.87%

attention
42.53%

intermediate
35.60%