gradients/bert/encoder/layer_2

output 17.96%

intermediate 19.24%

attention 62.80%