

gradients/bert/encoder/layer_3

output
17.90%

attention
62.71%

intermediate
19.38%

