

gradients/bert/encoder/layer_3

output
21.91%

attention
42.87%



intermediate
35.22%