Linear(in_features=1024, out_features=1024, bias=True)
[forward_compute_time=0.852,backward_compute_time=0.770,activation_size=16777216.0,parameter_size=4198400.0] ReLU() [forward_compute_time=0.089,backward_compute_time=0.086,activation_size=16777216.0,parameter_size=0.0] Linear(in_features=1024, out_features=1024, bias=True)
[forward_compute_time=0.786,backward_compute_time=1.456,activation_size=16777216.0,parameter_size=4198400.0] ReLU()
[forward_compute_time=0.077,backward_compute_time=0.083,activation_size=16777216.0,parameter_size=0.0] Linear(in_features=1024, out_features=64, bias=True)
[forward_compute_time=0.202,backward_compute_time=0.173,activation_size=1048576.0,parameter_size=262400.0] Input2 [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=0.0,parameter_size=0.0] Input1
[forward_compute_time=0.000,backward_compute_time=0.000,activation_size=0.0,parameter_size=0.0] EmbeddingBag(30000000, 64, mode=sum)
[forward_compute_time=0.229,backward_compute_time=0.255,activation_size=1048576.0,parameter_size=7680000000.0] ReLU()
[forward_compute_time=0.046,backward_compute_time=0.127,activation_size=1048576.0,parameter_size=0.0] Concat(1)
[forward_compute_time=0.000,backward_compute_time=0.000,activation_size=1048576.0,parameter_size=0.0] $View\\ [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=1048576.0,parameter_size=0.0]$ $Transpose (1,\,2) \\ [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=1048576.0,parameter_size=0.0]$ [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=0.0,parameter_size=0.0] __getitem__((slice(None, None, None), tensor([1]), tensor([0])))
[forward_compute_time=0.000,backward_compute_time=0.000,activation_size=0.0,parameter_size=0.0] Concat(1) [forward compute time=0.000,backward compute time=0.000,activation size=1048576.0,parameter size=0.0] Linear(in_features=65, out_features=128, bias=True)
[forward_compute_time=0.119,backward_compute_time=0.110,activation_size=2097152.0,parameter_size=33792.0] $ReLU() \\ [forward_compute_time=0.045,backward_compute_time=0.034,activation_size=2097152.0,parameter_size=0.0]$ Linear(in_features=128, out_features=512, bias=True)
[forward_compute_time=0.135,backward_compute_time=0.309,activation_size=8388608.0,parameter_size=264192.0] ReLU()
[forward_compute_time=0.056,backward_compute_time=0.055,activation_size=8388608.0,parameter_size=0.0] Linear(in_features=512, out_features=1024, bias=True)
[forward_compute_time=0.463,backward_compute_time=0.845,activation_size=16777216.0,parameter_size=2101248.0] ReLU() [forward_compute_time=0.076,backward_compute_time=0.086,activation_size=16777216.0,parameter_size=0.0] Linear(in_features=1024, out_features=1024, bias=True)
[forward_compute_time=0.784,backward_compute_time=1.451,activation_size=16777216.0,parameter_size=4198400.0] $ReLU() \\ [forward_compute_time=0.077,backward_compute_time=0.086,activation_size=16777216.0,parameter_size=0.0]$ Linear(in_features=1024, out_features=1024, bias=True)
[forward_compute_time=0.782,backward_compute_time=1.451,activation_size=16777216.0,parameter_size=4198400.0] ReLU()
[forward_compute_time=0.077,backward_compute_time=0.087,activation_size=16777216.0,parameter_size=0.0] Linear(in_features=1024, out_features=1024, bias=True)
[forward_compute_time=0.781,backward_compute_time=1.465,activation_size=16777216.0,parameter_size=4198400.0] ReLU() [forward_compute_time=0.077,backward_compute_time=0.085,activation_size=16777216.0,parameter_size=0.0] Linear(in_features=1024, out_features=1, bias=True)
[forward_compute_time=0.097,backward_compute_time=0.281,activation_size=16384.0,parameter_size=4100.0]

 $Input 0 \\ [forward_compute_time=0.000,backward_compute_time=0.000,activation_size=0.0,parameter_size=0.0]$

Sigmoid()
[forward_compute_time=0.050,backward_compute_time=0.055,activation_size=16384.0,parameter_size=0.0]