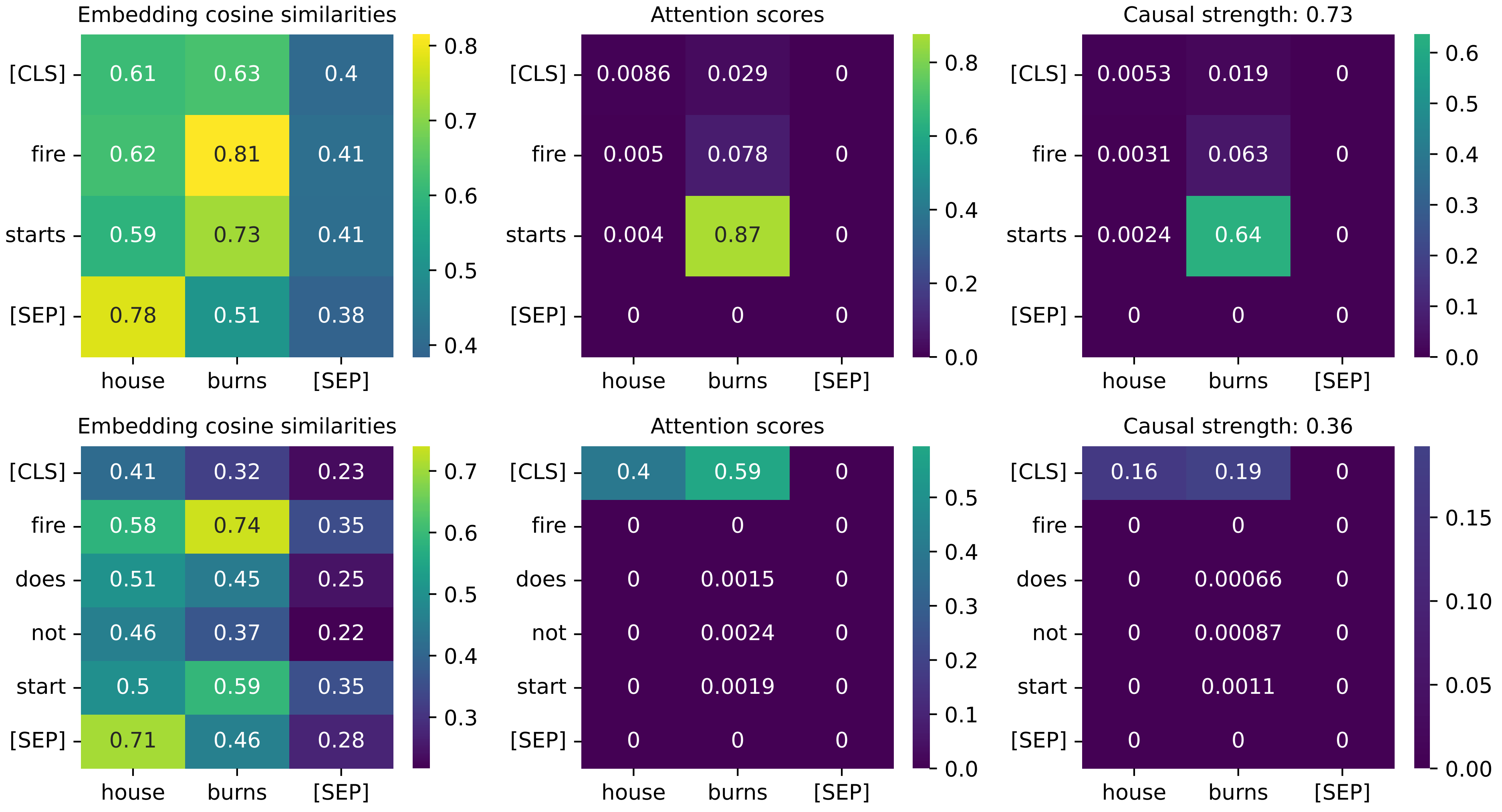
Embedding cosine similarities
[CLS]
fire
starts
[SEP]
0.61
0.63
0.4
0.62
0.81
0.41
0.59
0.73
0.41
0.78
0.51
0.38
house
burns
[SEP]
0.8
0.7
0.6
0.5
0.4
Attention scores
[CLS]
fire
starts
[SEP]
0.0086
0.029
0
0.005
0.078
0
0.004
0.87
0
0
0
0
house
burns
[SEP]
0.8
0.6
0.4
0.2
0.0
Causal strength: 0.73
[CLS]
fire
starts
[SEP]
0.0053
0.019
0
0.0031
0.063
0
0.0024
0.64
0
0
0
0
house
burns
[SEP]
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Embedding cosine similarities
[CLS]
fire
does
not
start
[SEP]
0.41
0.32
0.23
0.58
0.74
0.35
0.51
0.45
0.25
0.46
0.37
0.22
0.5
0.59
0.35
0.71
0.46
0.28
house
burns
[SEP]
0.7
0.6
0.5
0.4
0.3
Attention scores
[CLS]
fire
does
not
start
[SEP]
0.4
0.59
0
0
0
0
0
0.0015
0
0
0.0024
0
0
0.0019
0
0
0
0
house
burns
[SEP]
0.5
0.4
0.3
0.2
0.1
0.0
Causal strength: 0.36
[CLS]
fire
does
not
start
[SEP]
0.16
0.19
0
0
0
0
0
0.00066
0
0
0.00087
0
0
0.0011
0
0
0
0
house
burns
[SEP]
0.15
0.10
0.05
0.00