

Target	Human rationales	R2A-generated attention
Location	0.5185	0.2371
Cleanliness	0.5948	0.3141
Service	0.5833	0.2871

Table 6: Avg. cosine distance to the oracle attention over the target training set. The R2A is trained on beer reviews with unlabeled hotel reviews.