BERT + FCs:
Loading image features from img_features/ResNet-152-imagenet.tsv
Loading navigation graphs for 61 scans
R2RBatch loaded with 14039 instructions, using splits: train
Loading image features from img_features/ResNet-152-imagenet.tsv
Loading navigation graphs for 56 scans
R2RBatch loaded with 1021 instructions, using splits: val_seen
Loading image features from img_features/ResNet-152-imagenet.tsv
Loading navigation graphs for 11 scans
R2RBatch loaded with 2349 instructions, using splits: val_unseen
Build BERT model ...
Downloading: 100% 420M/420M [00:06<00:00, 68.6MB/s]
Some weights of the model checkpoint at bert-base-uncased were not used when initializing BertModel: ['cls.predictions.bias',
'cls.predictions.transform.LayerNorm.weight', 'cls.seq_relationship.weight', 'cls.predictions.transform.dense.bias', 'cls.predictions.decoder.weight',
'cls.predictions.transform.dense.weight', 'cls.predictions.transform.LayerNorm.bias', 'cls.seq_relationship.bias']
- This IS expected if you are initializing BertModel from the checkpoint of a model trained on another task or with another architecture (e.g. initializing a
BertForSequenceClassification model from a BertForPreTraining model).
- This IS NOT expected if you are initializing BertModel from the checkpoint of a model that you expect to be exactly identical (initializing a
BertForSequenceClassification model from a BertForSequenceClassification model).
Training with teacher feedback
100% 2000/2000 [15:21<00:00, 2.17it/s]
15m 50s (- 380m 14s) (2000 4%) train loss: 1.0932, val_seen loss: 1.0495, success_rate: 0.109, val_unseen loss: 1.2455, success_rate: 0.047
100% 2000/2000 [15:17<00:00, 2.18it/s]
31m 46s (- 365m 26s) (4000 8%) train loss: 0.9158, val_seen loss: 0.9989, success_rate: 0.090, val_unseen loss: 1.3249, success_rate: 0.065
100% 2000/2000 [15:26<00:00, 2.16it/s]
47m 50s (- 350m 53s) (6000 12%) train loss: 0.7975, val_seen loss: 0.9569, success_rate: 0.102, val_unseen loss: 1.4494, success_rate: 0.055
100% 2000/2000 [15:26<00:00, 2.16it/s]
63m 55s (- 335m 36s) (8000 16%) train loss: 0.7083, val_seen loss: 0.9312, success_rate: 0.105, val_unseen loss: 1.5411, success_rate: 0.055
100% 2000/2000 [15:19<00:00, 2.17it/s]
79m 53s (- 319m 34s) (10000 20%) train loss: 0.6404, val_seen loss: 0.9277, success_rate: 0.114, val_unseen loss: 1.6564, success_rate: 0.059
100% 2000/2000 [15:17<00:00, 2.18it/s]
95m 46s (- 303m 18s) (12000 24%) train loss: 0.5876, val_seen loss: 0.9145, success_rate: 0.106, val_unseen loss: 1.7742, success_rate: 0.063
100% 2000/2000 [15:19<00:00, 2.17it/s]
111m 43s (- 287m 17s) (14000 28%) train loss: 0.5458, val_seen loss: 0.9299, success_rate: 0.106, val_unseen loss: 1.8774, success_rate: 0.059
100% 2000/2000 [15:13<00:00, 2.19it/s]
127m 29s (- 270m 55s) (16000 32%) train loss: 0.5118, val_seen loss: 0.9455, success_rate: 0.114, val_unseen loss: 2.0063, success_rate: 0.053
100% 2000/2000 [15:12<00:00, 2.19it/s]
143m 18s (- 254m 46s) (18000 36%) train loss: 0.4834, val_seen loss: 0.9463, success_rate: 0.109, val_unseen loss: 2.0782, success_rate: 0.059
100% 2000/2000 [15:12<00:00, 2.19it/s]
159m 3s (- 238m 35s) (20000 40%) train loss: 0.4593, val_seen loss: 0.9749, success_rate: 0.130, val_unseen loss: 2.2032, success_rate: 0.052
100% 2000/2000 [15:12<00:00, 2.19it/s]
174m 48s (- 222m 29s) (22000 44%) train loss: 0.4389, val_seen loss: 0.9727, success_rate: 0.133, val_unseen loss: 2.2711, success_rate: 0.062
100% 2000/2000 [15:08<00:00, 2.20it/s]
190m 31s (- 206m 24s) (24000 48%) train loss: 0.4214, val_seen loss: 1.0047, success_rate: 0.142, val_unseen loss: 2.3905, success_rate: 0.053
100% 2000/2000 [15:12<00:00, 2.19it/s]
206m 16s (- 190m 24s) (26000 52%) train loss: 0.4053, val_seen loss: 1.0032, success_rate: 0.147, val_unseen loss: 2.4807, success_rate: 0.046
100% 2000/2000 [15:13<00:00, 2.19it/s]
222m 2s (- 174m 27s) (28000 56%) train loss: 0.3919, val_seen loss: 1.0090, success_rate: 0.140, val_unseen loss: 2.5659, success_rate: 0.050
100% 2000/2000 [15:14<00:00, 2.19it/s]
237m 49s (- 158m 32s) (30000 60%) train loss: 0.3799, val_seen loss: 1.0210, success_rate: 0.159, val_unseen loss: 2.5909, success_rate: 0.052
100% 2000/2000 [15:14<00:00, 2.19it/s]
253m 36s (- 142m 39s) (32000 64%) train loss: 0.3694, val_seen loss: 1.0575, success_rate: 0.170, val_unseen loss: 2.7759, success_rate: 0.052
100% 2000/2000 [15:13<00:00, 2.19it/s]
269m 22s (- 126m 46s) (34000 68%) train loss: 0.3596, val_seen loss: 1.0711, success_rate: 0.154, val_unseen loss: 2.8389, success_rate: 0.052
100% 2000/2000 [15:13<00:00, 2.19it/s]
285m 9s (- 110m 53s) (36000 72%) train loss: 0.3505, val_seen loss: 1.1213, success_rate: 0.157, val_unseen loss: 2.8602, success_rate: 0.059
100% 2000/2000 [15:22<00:00, 2.17it/s]
301m 4s (- 95m 4s) (38000 76%) train loss: 0.3423, val_seen loss: 1.1694, success_rate: 0.156, val_unseen loss: 3.0210, success_rate: 0.052
100% 2000/2000 [15:20<00:00, 2.17it/s]
316m 57s (- 79m 14s) (40000 80%) train loss: 0.3352, val_seen loss: 1.1378, success_rate: 0.164, val_unseen loss: 3.0670, success_rate: 0.054
100% 2000/2000 [15:19<00:00, 2.17it/s]
332m 50s (- 63m 23s) (42000 84%) train loss: 0.3285, val_seen loss: 1.1768, success_rate: 0.175, val_unseen loss: 3.1965, success_rate: 0.054
100% 2000/2000 [15:08<00:00, 2.20it/s]
348m 31s (- 47m 31s) (44000 88%) train loss: 0.3220, val_seen loss: 1.1715, success_rate: 0.175, val_unseen loss: 3.2078, success_rate: 0.049
100% 2000/2000 [15:09<00:00, 2.20it/s]
364m 14s (- 31m 40s) (46000 92%) train loss: 0.3165, val_seen loss: 1.2053, success_rate: 0.192, val_unseen loss: 3.3578, success_rate: 0.052
100% 2000/2000 [15:15<00:00, 2.18it/s]
379m 59s (- 15m 49s) (48000 96%) train loss: 0.3111, val_seen loss: 1.2017, success_rate: 0.186, val_unseen loss: 3.4525, success_rate: 0.049
100% 2000/2000 [15:15<00:00, 2.18it/s]
395m 45s (- 0m 0s) (50000 100%) train loss: 0.3065, val_seen loss: 1.2836, success_rate: 0.180, val_unseen loss: 3.5629, success_rate: 0.055

Baseline:

Training with teacher feedback

100% 2000/2000 [06:52<00:00, 4.84it/s]

7m 4s (- 169m 45s) (2000 4%) train loss: 0.7944, val_seen loss: 0.7405, success_rate: 0.228, val_unseen loss: 0.8717, success_rate: 0.192

100% 2000/2000 [06:50<00:00, 4.87it/s]

14m 7s (- 162m 25s) (4000 8%) train loss: 0.5719, val_seen loss: 0.7144, success_rate: 0.223, val_unseen loss: 0.9624, success_rate: 0.201

100% 2000/2000 [06:49<00:00, 4.88it/s]

21m 9s (- 155m 12s) (6000 12%) train loss: 0.4495, val_seen loss: 0.7572, success_rate: 0.253, val_unseen loss: 1.1093, success_rate: 0.199

100% 2000/2000 [06:50<00:00, 4.88it/s]

28m 12s (- 148m 5s) (8000 16%) train loss: 0.3604, val_seen loss: 0.8764, success_rate: 0.283, val_unseen loss: 1.3047, success_rate: 0.209

100% 2000/2000 [06:48<00:00, 4.89it/s]

35m 13s (- 140m 55s) (10000 20%) train loss: 0.2915, val_seen loss: 0.9217, success_rate: 0.268, val_unseen loss: 1.4291, success_rate: 0.193

100% 2000/2000 [06:49<00:00, 4.89it/s]

42m 15s (- 133m 48s) (12000 24%) train loss: 0.2357, val_seen loss: 0.9889, success_rate: 0.271, val_unseen loss: 1.5868, success_rate: 0.181

100% 2000/2000 [06:49<00:00, 4.88it/s]

49m 16s (- 126m 42s) (14000 28%) train loss: 0.1911, val_seen loss: 1.0352, success_rate: 0.265, val_unseen loss: 1.7409, success_rate: 0.183

100% 2000/2000 [06:50<00:00, 4.87it/s]

56m 18s (- 119m 38s) (16000 32%) train loss: 0.1545, val_seen loss: 1.1496, success_rate: 0.274, val_unseen loss: 1.9051, success_rate: 0.181

100% 2000/2000 [06:50<00:00, 4.87it/s]

63m 19s (- 112m 34s) (18000 36%) train loss: 0.1254, val_seen loss: 1.2088, success_rate: 0.268, val_unseen loss: 2.0140, success_rate: 0.169

100% 2000/2000 [06:49<00:00, 4.89it/s]

70m 19s (- 105m 29s) (20000 40%) train loss: 0.1014, val_seen loss: 1.3504, success_rate: 0.261, val_unseen loss: 2.1931, success_rate: 0.176

100% 2000/2000 [06:50<00:00, 4.88it/s]

77m 20s (- 98m 26s) (22000 44%) train loss: 0.0829, val_seen loss: 1.3900, success_rate: 0.257, val_unseen loss: 2.3038, success_rate: 0.179

100% 2000/2000 [06:49<00:00, 4.89it/s]

84m 21s (- 91m 22s) (24000 48%) train loss: 0.0689, val_seen loss: 1.4400, success_rate: 0.248, val_unseen loss: 2.3843, success_rate: 0.168

100% 2000/2000 [06:49<00:00, 4.88it/s]

91m 22s (- 84m 20s) (26000 52%) train loss: 0.0578, val_seen loss: 1.5511, success_rate: 0.249, val_unseen loss: 2.4270, success_rate: 0.182

100% 2000/2000 [06:50<00:00, 4.88it/s]

98m 23s (- 77m 18s) (28000 56%) train loss: 0.0498, val_seen loss: 1.6130, success_rate: 0.251, val_unseen loss: 2.5347, success_rate: 0.181

100% 2000/2000 [06:48<00:00, 4.90it/s]

105m 22s (- 70m 15s) (30000 60%) train loss: 0.0435, val_seen loss: 1.6383, success_rate: 0.248, val_unseen loss: 2.5662, success_rate: 0.192

100% 2000/2000 [06:48<00:00, 4.90it/s]

112m 22s (- 63m 12s) (32000 64%) train loss: 0.0387, val_seen loss: 1.6864, success_rate: 0.248, val_unseen loss: 2.6570, success_rate: 0.184

100% 2000/2000 [06:48<00:00, 4.90it/s]

119m 21s (- 56m 10s) (34000 68%) train loss: 0.0351, val_seen loss: 1.6681, success_rate: 0.240, val_unseen loss: 2.6061, success_rate: 0.165

100% 2000/2000 [06:47<00:00, 4.91it/s]

126m 20s (- 49m 7s) (36000 72%) train loss: 0.0318, val_seen loss: 1.6837, success_rate: 0.255, val_unseen loss: 2.6921, success_rate: 0.172

100% 2000/2000 [06:47<00:00, 4.91it/s]

133m 18s (- 42m 5s) (38000 76%) train loss: 0.0291, val_seen loss: 1.7271, success_rate: 0.239, val_unseen loss: 2.7279, success_rate: 0.173

100% 2000/2000 [06:48<00:00, 4.89it/s]

140m 18s (- 35m 4s) (40000 80%) train loss: 0.0271, val_seen loss: 1.7269, success_rate: 0.258, val_unseen loss: 2.7794, success_rate: 0.178

100% 2000/2000 [06:48<00:00, 4.90it/s]

147m 17s (- 28m 3s) (42000 84%) train loss: 0.0259, val_seen loss: 1.8087, success_rate: 0.248, val_unseen loss: 2.8642, success_rate: 0.177

100% 2000/2000 [06:48<00:00, 4.90it/s]

154m 16s (- 21m 2s) (44000 88%) train loss: 0.0242, val_seen loss: 1.8152, success_rate: 0.239, val_unseen loss: 2.8433, success_rate: 0.180

100% 2000/2000 [06:47<00:00, 4.90it/s]

161m 16s (- 14m 1s) (46000 92%) train loss: 0.0232, val_seen loss: 1.7874, success_rate: 0.251, val_unseen loss: 2.7988, success_rate: 0.182

100% 2000/2000 [06:47<00:00, 4.90it/s]

168m 15s (- 7m 0s) (48000 96%) train loss: 0.0217, val_seen loss: 1.8842, success_rate: 0.255, val_unseen loss: 2.9635, success_rate: 0.183

100% 2000/2000 [06:48<00:00, 4.89it/s]

175m 15s (- 0m 0s) (50000 100%) train loss: 0.0213, val_seen loss: 1.9397, success_rate: 0.252, val_unseen loss: 2.8621, success_rate: 0.189

BERT + Decoder-LSTM:

Build BERT+LSTM model ...
Some weights of the model checkpoint at bert-base-uncased were not used when initializing BertModel: ['cls.predictions.bias',
'cls.predictions.transform.dense.bias', 'cls.predictions.transform.LayerNorm.weight', 'cls.predictions.transform.LayerNorm.bias',
'cls.predictions.decoder.weight', 'cls.seq_relationship.bias', 'cls.seq_relationship.weight', 'cls.predictions.transform.dense.weight']
- This IS expected if you are initializing BertModel from the checkpoint of a model trained on another task or with another architecture (e.g. initializing a
BertForSequenceClassification model from a BertForPreTraining model).
- This IS NOT expected if you are initializing BertModel from the checkpoint of a model that you expect to be exactly identical (initializing a
BertForSequenceClassification model from a BertForSequenceClassification model).
Training with teacher feedback
100% 2000/2000 [16:34<00:00, 2.01it/s]
17m 5s (- 410m 6s) (2000 4%) train loss: 0.8692, val_seen loss: 0.8057, success_rate: 0.189, val_unseen loss: 0.8718, success_rate: 0.164
100% 2000/2000 [16:28<00:00, 2.02it/s]
34m 9s (- 392m 46s) (4000 8%) train loss: 0.7247, val_seen loss: 0.7712, success_rate: 0.175, val_unseen loss: 0.9036, success_rate: 0.153
100% 2000/2000 [16:28<00:00, 2.02it/s]
51m 14s (- 375m 43s) (6000 12%) train loss: 0.6448, val_seen loss: 0.7620, success_rate: 0.200, val_unseen loss: 0.9413, success_rate: 0.164
100% 2000/2000 [16:28<00:00, 2.02it/s]
68m 18s (- 358m 35s) (8000 16%) train loss: 0.5783, val_seen loss: 0.7653, success_rate: 0.192, val_unseen loss: 0.9944, success_rate: 0.145
100% 2000/2000 [16:33<00:00, 2.01it/s]
85m 26s (- 341m 44s) (10000 20%) train loss: 0.5254, val_seen loss: 0.7847, success_rate: 0.205, val_unseen loss: 1.0798, success_rate: 0.149
100% 2000/2000 [16:33<00:00, 2.01it/s]
102m 37s (- 324m 58s) (12000 24%) train loss: 0.4780, val_seen loss: 0.8215, success_rate: 0.198, val_unseen loss: 1.1517, success_rate: 0.149
100% 2000/2000 [16:33<00:00, 2.01it/s]
119m 49s (- 308m 6s) (14000 28%) train loss: 0.4376, val_seen loss: 0.8533, success_rate: 0.217, val_unseen loss: 1.2561, success_rate: 0.148
100% 2000/2000 [16:34<00:00, 2.01it/s]
136m 58s (- 291m 4s) (16000 32%) train loss: 0.4014, val_seen loss: 0.9070, success_rate: 0.194, val_unseen loss: 1.3074, success_rate: 0.144
100% 2000/2000 [16:34<00:00, 2.01it/s]
154m 8s (- 274m 0s) (18000 36%) train loss: 0.3685, val_seen loss: 0.9366, success_rate: 0.193, val_unseen loss: 1.4062, success_rate: 0.146
100% 2000/2000 [16:35<00:00, 2.01it/s]
171m 18s (- 256m 57s) (20000 40%) train loss: 0.3396, val_seen loss: 0.9378, success_rate: 0.189, val_unseen loss: 1.4617, success_rate: 0.148
100% 2000/2000 [16:36<00:00, 2.01it/s]
188m 29s (- 239m 53s) (22000 44%) train loss: 0.3140, val_seen loss: 0.9846, success_rate: 0.194, val_unseen loss: 1.5109, success_rate: 0.150
100% 2000/2000 [16:34<00:00, 2.01it/s]
205m 38s (- 222m 46s) (24000 48%) train loss: 0.2896, val_seen loss: 1.0210, success_rate: 0.195, val_unseen loss: 1.5967, success_rate: 0.154
100% 2000/2000 [16:35<00:00, 2.01it/s]
222m 49s (- 205m 40s) (26000 52%) train loss: 0.2687, val_seen loss: 1.0431, success_rate: 0.190, val_unseen loss: 1.6876, success_rate: 0.159
100% 2000/2000 [16:34<00:00, 2.01it/s]
239m 59s (- 188m 33s) (28000 56%) train loss: 0.2493, val_seen loss: 1.0721, success_rate: 0.201, val_unseen loss: 1.7084, success_rate: 0.154
100% 2000/2000 [16:29<00:00, 2.02it/s]
257m 3s (- 171m 22s) (30000 60%) train loss: 0.2322, val_seen loss: 1.0951, success_rate: 0.204, val_unseen loss: 1.7511, success_rate: 0.163
100% 2000/2000 [16:34<00:00, 2.01it/s]
274m 12s (- 154m 14s) (32000 64%) train loss: 0.2162, val_seen loss: 1.1134, success_rate: 0.174, val_unseen loss: 1.8077, success_rate: 0.159
100% 2000/2000 [16:33<00:00, 2.01it/s]
291m 20s (- 137m 5s) (34000 68%) train loss: 0.2023, val_seen loss: 1.1353, success_rate: 0.189, val_unseen loss: 1.8403, success_rate: 0.157
100% 2000/2000 [16:35<00:00, 2.01it/s]
308m 29s (- 119m 58s) (36000 72%) train loss: 0.1891, val_seen loss: 1.1836, success_rate: 0.202, val_unseen loss: 1.9077, success_rate: 0.149
100% 2000/2000 [16:32<00:00, 2.02it/s]
325m 36s (- 102m 49s) (38000 76%) train loss: 0.1775, val_seen loss: 1.2095, success_rate: 0.197, val_unseen loss: 1.9522, success_rate: 0.146
100% 2000/2000 [16:30<00:00, 2.02it/s]
342m 42s (- 85m 40s) (40000 80%) train loss: 0.1674, val_seen loss: 1.2402, success_rate: 0.208, val_unseen loss: 2.0382, success_rate: 0.152
100% 2000/2000 [16:32<00:00, 2.01it/s]
359m 49s (- 68m 32s) (42000 84%) train loss: 0.1576, val_seen loss: 1.2254, success_rate: 0.206, val_unseen loss: 2.0427, success_rate: 0.166
100% 2000/2000 [16:29<00:00, 2.02it/s]
376m 53s (- 51m 23s) (44000 88%) train loss: 0.1492, val_seen loss: 1.2853, success_rate: 0.198, val_unseen loss: 2.1055, success_rate: 0.151
100% 2000/2000 [16:29<00:00, 2.02it/s]
393m 57s (- 34m 15s) (46000 92%) train loss: 0.1410, val_seen loss: 1.2783, success_rate: 0.199, val_unseen loss: 2.0986, success_rate: 0.147
100% 2000/2000 [16:26<00:00, 2.03it/s]
410m 57s (- 17m 7s) (48000 96%) train loss: 0.1340, val_seen loss: 1.3548, success_rate: 0.196, val_unseen loss: 2.1691, success_rate: 0.147
100% 2000/2000 [16:25<00:00, 2.03it/s]
427m 57s (- 0m 0s) (50000 100%) train loss: 0.1280, val_seen loss: 1.3271, success_rate: 0.193, val_unseen loss: 2.1897, success_rate: 0.150