

A Hyperparameters

Hyperparameter	Value
Number of units in the utterance-level RNN	150
Number of units in the context-level RNN	150
Number of units in the decoder (if used)	150
Optimizer	Adam
Learning rate	0.001
Gradient clipping	5.0
Dropout	0.5
Batch size	64
Number of epochs	15
Vocabulary size	1000

Table 6: Training and model hyperparameters

	NUR	NUG	DAP	BSP
No	24.20	8.66	15.11	5.30
NUR	–	12.43	18.31	7.53
NUG	35.93	–	19.50	8.23
MUR	31.57	10.71	18.82	5.90
InI	26.63	10.84	15.42	6.40

Table 9: Results with 5% of the data

B Entities in BS vector

Domain	Entities	# values
Taxi	'leaveAt', 'destination', 'departure', 'arriveBy'	839
Restaurant	'time', 'day', 'people', 'food', 'pricerange', 'area'	269
Hospital	'department'	52
Hotel	'stay', 'day', 'people', 'area', 'parking', 'pricerange', 'stars', 'internet', 'type'	143
Attraction	'type', 'area'	67
Train	'people', 'ticket', 'leaveAt', 'destination', 'day', 'arriveBy', 'departure'	414

Table 7: Values used in the belief state

C Low resource setting results

	NUR	NUG	DAP	BSP
No	17.28	7.00	17.04	4.80
NUR	–	7.55	17.46	7.10
NUG	32.95	–	19.67	8.37
MUR	28.89	11.84	17.98	5.81
InI	26.00	10.55	14.31	6.31

Table 8: Results with 2% of the data

	NUR	NUG	DAP	BSP
No	47.57	13.36	25.12	7.98
NUR	–	15.30	28.81	12.52
NUG	51.37	–	29.29	11.28
MUR	46.06	14.56	29.77	11.59
InI	47.52	14.87	30.19	11.33

Table 10: Results with 50% of the data