

A Experimental Details

A.1 Hyper-parameters

Hyper-parameter	Value
learning rate	3e-5
weight decay	0.01
batch size	48
epoch	2
optimiser	Adam
Adam ϵ	1e-6
Adam (β_1, β_2)	(0.9, 0.999)
warmup ratio	10%
max sequence length	200
max question length	100
max answer length	30
number of passages	5
dropout	0.0
pretrained model	albert-large-v2
number of parameters	18M
device	Nvidia Titan X

Table 5: Hyper-parameters for reader model training.

Hyper-parameter	Value
learning rate	1e-3
batch size	32
epoch	16
optimiser	SGD
max number of steps	240
step cost c	0.1
discount factor γ	0.9
number of passages	30

Table 6: Hyper-parameters for scheduler model RL training.