

A Appendices: Hyperparameters

Hyperparams	BERT		RoBERTa	
	MRC	Reflection	MRC	Reflection
Dropout		0.1		0.1
Attention dropout		0.1		0.1
Max sequence length		512		512
Learning Rate	3e-5	5e-6	2.2e-5	5e-6
Batch Size		24		48
Weight Decay		0.01		0.01
Epochs	1	2	1	2
Learning Rate Decay		Linear		Linear
Warmup ratio		0.1		0.06
Gradient Clipping		1.0		-
Adam ϵ		1e-8		1e-8
Adam β_1		0.9		0.9
Adam β_2		0.999		0.999

Table 1: Training hyperparameters of MRC model and Reflection model.