

Train model	BERT-base	RoBERTa-base	DeBERTa-v3-base
Learning rate	2e-5/3e-5	2e-5/2e-5	2e-5/2e-5
Train epochs	3/2	3/3	3/2
Batch size	32/32	32/32	64/64
Weight decay	0.01/0.1	0.1/0.01	0.0/0.0
Train data	MNLI / SNLI		
Test accuracy	84.04/83.93 (-m/-mm) 89.44	87.84/87.62 (-m/-mm) 91.03	90.21/90.73 (-m/-mm) 93.33
Train data	aug-MNLI / aug-SNLI		
Test accuracy	82.72/83.12 (-m/-mm) 90.98	87.63/87.53 (-m/-mm) 91.98	90.37/90.44 (-m/-mm) 93.40

Train model	BERT-large	RoBERTa-large	DeBERTa-v3-large
Learning rate	2e-5/3e-5	6e-6/6e-6	6e-6/5e-6
Train epochs	2/2	2/2	2/2
Batch size	32/32	64/64	32/32
Weight decay	0.01/0.1	0.0/0.0	0.0/0.0
Train data	MNLI / SNLI		
Test accuracy	86.24/86.51 (-m/-mm) 91.85	90.34/90.13 (-m/-mm) 92.66	91.18/91.10 (-m/-mm) 93.92
Train data	aug-MNLI / aug-SNLI		
Test accuracy	85.77/86.01(-m/-mm) 91.53	90.80/90.53 (-m/-mm) 93.18	91.69/91.67 (-m/-mm) 93.94