

Model & Method	# Trainable Parameters	MNLI	SST-2	MRPC	CoLA	QNLI	QQP	RTE	STS-B	Avg.
RoB _{base} (FT)*	125.0M	87.6	94.8	90.2	63.6	92.8	91.9	78.7	91.2	86.4
RoB _{base} (BitFit)*	0.1M	84.7	93.7	92.7	62.0	91.8	84.0	81.5	90.8	85.2
RoB _{base} (Adpt [†])*	0.3M	87.1 _{±0}	94.2 _{±1}	88.5 _{±1}	60.8 _{±4}	93.1 _{±1}	90.2 _{±0}	71.5 _{±27}	89.7 _{±3}	84.4
RoB _{base} (Adpt [†])*	0.9M	87.3 _{±1}	94.7 _{±3}	88.4 _{±1}	62.6 _{±9}	93.0 _{±2}	90.6 _{±0}	75.9 _{±22}	90.3 _{±1}	85.4
RoB _{base} (LoRA)	0.3M	87.5 _{±3}	95.1 _{±2}	89.7 _{±7}	63.4 _{±12}	93.3 _{±3}	90.8 _{±1}	86.6 _{±7}	91.5 _{±2}	87.2
RoB _{large} (FT)*	355.0M	90.2	96.4	90.9	68.0	94.7	92.2	86.6	92.4	88.9
RoB _{large} (LoRA)	0.8M	90.6 _{±2}	96.2 _{±5}	90.9 _{±12}	68.2 _{±19}	94.9 _{±3}	91.6 _{±1}	87.4 _{±25}	92.6 _{±2}	89.0
RoB _{large} (Adpt [†])*	3.0M	90.2 _{±3}	96.1 _{±3}	90.2 _{±7}	68.3 _{±10}	94.8 _{±2}	91.9 _{±1}	83.8 _{±29}	92.1 _{±7}	88.4
RoB _{large} (Adpt [†])*	0.8M	90.5 _{±3}	96.6 _{±2}	89.7 _{±12}	67.8 _{±25}	94.8 _{±3}	91.7 _{±2}	80.1 _{±29}	91.9 _{±4}	87.9
RoB _{large} (Adpt [†])*	6.0M	89.9 _{±5}	96.2 _{±3}	88.7 _{±12}	66.5 _{±44}	94.7 _{±2}	92.1 _{±1}	83.4 _{±11}	91.0 _{±17}	87.8
RoB _{large} (Adpt [†])*	0.8M	90.3 _{±3}	96.3 _{±5}	87.7 _{±17}	66.3 _{±20}	94.7 _{±2}	91.5 _{±1}	72.9 _{±29}	91.5 _{±5}	86.4
RoB _{large} (LoRA)*	0.8M	90.6 _{±2}	96.2 _{±5}	90.2 _{±10}	68.2 _{±19}	94.8 _{±3}	91.6 _{±2}	85.2 _{±11}	92.3 _{±5}	88.6
DeB _{xxl} (FT)*	1500.0M	91.8	97.2	92.0	72.0	96.0	92.7	93.9	92.9	91.1
DeB _{xxl} (LoRA)	4.7M	91.9 _{±2}	96.9 _{±2}	92.6 _{±6}	72.4 _{±11}	96.0 _{±1}	92.9 _{±1}	94.9 _{±4}	93.0 _{±2}	91.3

Table 2: RoBERTa_{base}, RoBERTa_{large}, and DeBERTa_{xxl} with different adaptation methods on the GLUE benchmark. We report the overall (matched and mismatched) accuracy for MNLI, Matthew's correlation for CoLA, Pearson correlation for STS-B, and accuracy for other tasks. Higher is better for all metrics. * indicates numbers published in prior works. † indicates runs configured in a setup similar to [Houlsby et al. \(2019\)](#) for a fair comparison.

benchmark

7b pretrain accuracy: **0.5195**

```

root@e8e99ab71e21: ~
Evaluating: 99%|*****| 863/872 [03:05<00:02, 3.93it/s]
Probabilities for choices: {'1': 0.6875, '0': 0.312255859375}
Predicted Label: 1, True Label: 0
Evaluating: 99%|*****| 864/872 [03:05<00:02, 3.87it/s]
Probabilities for choices: {'1': 0.56005859375, '0': 0.439697265625}
Predicted Label: 1, True Label: 0
Probabilities for choices: {'1': 0.68603515625, '0': 0.31396484375}*****
Predicted Label: 1, True Label: 1
Evaluating: 99%|*****| 866/872 [03:06<00:01, 4.32it/s]
Probabilities for choices: {'1': 0.6279296875, '0': 0.3720703125}
Predicted Label: 1, True Label: 1
Evaluating: 99%|*****| 867/872 [03:06<00:01, 4.19it/s]
Probabilities for choices: {'1': 0.63330078125, '0': 0.36669921875}
Predicted Label: 1, True Label: 0
Evaluating: 100%|*****| 868/872 [03:06<00:00, 4.39it/s]
Probabilities for choices: {'1': 0.642578125, '0': 0.357666015625}
Predicted Label: 1, True Label: 1
Evaluating: 100%|*****| 869/872 [03:06<00:00, 4.56it/s]
Probabilities for choices: {'1': 0.59619140625, '0': 0.40380859375}
Predicted Label: 1, True Label: 0
Evaluating: 100%|*****| 870/872 [03:06<00:00, 4.68it/s]
Probabilities for choices: {'1': 0.6318359375, '0': 0.3681640625}
Predicted Label: 1, True Label: 0
Evaluating: 100%|*****| 871/872 [03:07<00:00, 4.76it/s]
Probabilities for choices: {'1': 0.67041015625, '0': 0.329345703125}
Predicted Label: 1, True Label: 1
Evaluating: 100%|*****| 872/872 [03:07<00:00, 4.65it/s]
Evaluation completed. Accuracy: 0.5195 (453/872)
Logits saved!
(base) root@e8e99ab71e21:~/Workspace/SF_CPT/SST-2#

```

13b pretrain accuracy: **0.8337**

```

root@e8e99ab71e21: ~
Evaluating: 99%|*****| 863/872 [06:29<00:04, 1.99it/s]
Probabilities for choices: {'1': 0.513671875, '0': 0.486328125}
Predicted Label: 1, True Label: 0
Evaluating: 99%|*****| 864/872 [06:30<00:04, 1.96it/s]
Probabilities for choices: {'1': 0.178955078125, '0': 0.82080078125}
Predicted Label: 0, True Label: 0
Probabilities for choices: {'1': 0.529296875, '0': 0.470703125}*****
Predicted Label: 1, True Label: 1
Evaluating: 99%|*****| 866/872 [06:31<00:02, 2.12it/s]
Probabilities for choices: {'1': 0.66552734375, '0': 0.33447265625}
Predicted Label: 1, True Label: 1
Evaluating: 99%|*****| 867/872 [06:31<00:02, 2.06it/s]
Probabilities for choices: {'1': 0.42822265625, '0': 0.57177734375}
Predicted Label: 0, True Label: 0
Evaluating: 100%|*****| 868/872 [06:32<00:01, 2.14it/s]
Probabilities for choices: {'1': 0.544921875, '0': 0.455078125}
Predicted Label: 1, True Label: 1
Evaluating: 100%|*****| 869/872 [06:32<00:01, 2.21it/s]
Probabilities for choices: {'1': 0.346923828125, '0': 0.653076171875}
Predicted Label: 0, True Label: 0
Evaluating: 100%|*****| 870/872 [06:33<00:00, 2.26it/s]
Probabilities for choices: {'1': 0.361083984375, '0': 0.638916015625}
Predicted Label: 0, True Label: 0
Evaluating: 100%|*****| 871/872 [06:33<00:00, 2.28it/s]
Probabilities for choices: {'1': 0.55810546875, '0': 0.44189453125}
Predicted Label: 1, True Label: 1
Evaluating: 100%|*****| 872/872 [06:33<00:00, 2.21it/s]
Evaluation completed. Accuracy: 0.8337 (727/872)
Logits saved!
(base) root@e8e99ab71e21:~/Workspace/SF_CPT/SST-2#

```

7b lora-tune accuracy: **0.9599**

```
root@e8e99ab71e21: ~/Work: x + v
Probabilities for choices: {'1': 0.0271759033203125, '0': 0.97265625}
Predicted Label: 0, True Label: 0
Evaluating: 99%|*****| 864/872 [03:23<00:02, 3.68it/s]
Probabilities for choices: {'1': 0.011871337890625, '0': 0.98828125}
Predicted Label: 0, True Label: 0
Probabilities for choices: {'1': 0.08038330078125, '0': 0.91943359375}*****
Predicted Label: 0, True Label: 1
Evaluating: 99%|*****| 865/872 [03:23<00:01, 3.92it/s]
Probabilities for choices: {'1': 0.90869140625, '0': 0.09136962890625}
Predicted Label: 1, True Label: 1
Evaluating: 99%|*****| 866/872 [03:23<00:01, 4.10it/s]
Probabilities for choices: {'1': 0.0323486328125, '0': 0.9677734375}
Predicted Label: 0, True Label: 0
Evaluating: 100%|*****| 867/872 [03:24<00:01, 3.95it/s]
Probabilities for choices: {'1': 0.99267578125, '0': 0.007404327392578125}
Predicted Label: 1, True Label: 1
Evaluating: 100%|*****| 868/872 [03:24<00:00, 4.12it/s]
Probabilities for choices: {'1': 0.01384735107421875, '0': 0.986328125}
Predicted Label: 0, True Label: 0
Evaluating: 100%|*****| 869/872 [03:24<00:00, 4.26it/s]
Probabilities for choices: {'1': 0.011688232421875, '0': 0.98828125}
Predicted Label: 0, True Label: 0
Evaluating: 100%|*****| 870/872 [03:24<00:00, 4.36it/s]
Probabilities for choices: {'1': 0.9951171875, '0': 0.005100250244140625}
Predicted Label: 1, True Label: 1
Evaluating: 100%|*****| 871/872 [03:24<00:00, 4.43it/s]
Probabilities for choices: {'1': 0.98193359375, '0': 0.017852783203125}
Predicted Label: 1, True Label: 1
Evaluating: 100%|*****| 872/872 [03:25<00:00, 4.25it/s]
Evaluation completed. Accuracy: 0.9599 (837/872)
Logits saved!
Finished: SST-2/7b_tune_acc.py
(base) root@e8e99ab71e21:~/Workspace/SF_CPT#
```

13b lora-tune accuracy: **0.9587**

```
root@e8e99ab71e21: ~ x + v
Evaluating: 99%|*****| 863/872 [06:32<00:04, 1.93it/s]
Probabilities for choices: {'1': 0.1015625, '0': 0.8984375}
Predicted Label: 0, True Label: 0
Evaluating: 99%|*****| 864/872 [06:33<00:04, 1.90it/s]
Probabilities for choices: {'1': 0.041778564453125, '0': 0.9580078125}
Predicted Label: 0, True Label: 0
Probabilities for choices: {'1': 0.363037109375, '0': 0.63720703125}*****
Predicted Label: 0, True Label: 1
Evaluating: 99%|*****| 865/872 [06:33<00:03, 2.01it/s]
Probabilities for choices: {'1': 0.94189453125, '0': 0.058349609375}
Predicted Label: 1, True Label: 1
Evaluating: 99%|*****| 866/872 [06:34<00:02, 2.08it/s]
Probabilities for choices: {'1': 0.06097412109375, '0': 0.93896484375}
Predicted Label: 0, True Label: 0
Evaluating: 99%|*****| 867/872 [06:34<00:02, 2.01it/s]
Probabilities for choices: {'1': 0.98095703125, '0': 0.0191192626953125}
Predicted Label: 1, True Label: 1
Evaluating: 100%|*****| 868/872 [06:35<00:01, 2.08it/s]
Probabilities for choices: {'1': 0.1634521484375, '0': 0.83642578125}
Predicted Label: 0, True Label: 0
Evaluating: 100%|*****| 869/872 [06:35<00:01, 2.13it/s]
Probabilities for choices: {'1': 0.04742431640625, '0': 0.95263671875}
Predicted Label: 0, True Label: 0
Evaluating: 100%|*****| 870/872 [06:36<00:00, 2.17it/s]
Probabilities for choices: {'1': 0.98193359375, '0': 0.017852783203125}
Predicted Label: 1, True Label: 1
Evaluating: 100%|*****| 871/872 [06:36<00:00, 2.21it/s]
Probabilities for choices: {'1': 0.98193359375, '0': 0.017852783203125}
Predicted Label: 1, True Label: 1
Evaluating: 100%|*****| 872/872 [06:36<00:00, 2.20it/s]
Evaluation completed. Accuracy: 0.9587 (836/872)
Logits saved!
(base) root@e8e99ab71e21:~/Workspace/SF_CPT/SST-2#
```

Proxy tuning accuracy: **0.9622**

```
root@e8e99ab71e21: ~/Work: x + v
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.9903663396835327, '0': 0.009633652865886688}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.8459424376487732, '0': 0.1540575623512268}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.9700134992599487, '0': 0.029986506327986717}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.047074176371097565, '0': 0.9529258608818054}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.015306354500353336, '0': 0.9846936464309692}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.0029349515680223703, '0': 0.997065007686615}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.04922211170196533, '0': 0.9507778882980347}
Predicted Answer: 0, True Answer: 1
Probabilities for choices: {'1': 0.9190633296966553, '0': 0.08093671500682831}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.016657041385769844, '0': 0.9833430051803589}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.9897513389587402, '0': 0.01024860143661499}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.006145091727375984, '0': 0.993854820728302}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.004867417272180319, '0': 0.9951325058937073}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.9925393462181091, '0': 0.007460643537342548}
Predicted Answer: 1, True Answer: 1
Evaluating: 100% | 872/872 [00:00<00:00, 2027.64it/s]
Evaluation completed. Accuracy: 0.9622 (839/872)
(base) root@e8e99ab71e21:~/Workspace/SF_CPT/SST-2#
```

Consistent proxy tuning accuracy: **0.9679**

```
root@e8e99ab71e21: ~/Work: x + v
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.9996898174285889, '0': 0.00031015597051009536}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.9993833303451538, '0': 0.0006166296661831439}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.999562680721283, '0': 0.0004373344127088785}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.0017546144081279635, '0': 0.998245358467102}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.0018102109897881746, '0': 0.9981897473335266}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.00018522574100643396, '0': 0.9998148083686829}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.004609571769833565, '0': 0.9953904151916504}
Predicted Answer: 0, True Answer: 1
Probabilities for choices: {'1': 0.9674102663993835, '0': 0.03258974477648735}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.0007437028689309955, '0': 0.9992563128471375}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.999713122844696, '0': 0.0002868540759664029}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.0006771942134946585, '0': 0.999322772026062}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.0004173157794866711, '0': 0.9995827078819275}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.9998058676719666, '0': 0.00019411317771300673}
Predicted Answer: 1, True Answer: 1
Evaluating: 100% | 872/872 [00:00<00:00, 2261.08it/s]
Evaluation completed. Accuracy: 0.9679 (844/872)
(base) root@e8e99ab71e21:~/Workspace/SF_CPT/SST-2#
```

Gaussian Process tuning accuracy: **0.9667**

800 / 9000 (8.89%)

```
root@e8e99ab71e21: ~/Work
Probabilities for choices: {'1': 0.9997305274009705, '0': 0.0002694791473913938}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.999438464641571, '0': 0.0005614787223748863}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.9996485710144043, '0': 0.0003514382697176188}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.0013458604225888848, '0': 0.9986541271209717}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.0011695101857185364, '0': 0.9988304972648621}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.00011235327110625803, '0': 0.9998875856399536}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.005060331895947456, '0': 0.9949396848678589}
Predicted Answer: 0, True Answer: 1
Probabilities for choices: {'1': 0.6548947095870972, '0': 0.34510529041290283}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.0004955359036102891, '0': 0.9995044469833374}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.9997695088386536, '0': 0.00023050671734381467}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.00035697061684913933, '0': 0.9996429681777954}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.00026530236937105656, '0': 0.9997346997261047}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.9998313188552856, '0': 0.00016865275392774493}
Predicted Answer: 1, True Answer: 1
Evaluating: 100% | 872/872 [00:00<00:00, 2260.97it/s]
Evaluation completed. Accuracy: 0.9667 (843/872)
(base) root@e8e99ab71e21:~/Workspace/SF_CPT/SST-2# python 7b_tune_acc.py
```

Gaussian Process with filter tuning accuracy: **0.9656**

Config: input_threshold=0.115, output_threshold=0.8

173 / 9000 (1.92%)

```
root@e8e99ab71e21: ~/Work
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.9997305274009705, '0': 0.0002694791473913938}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.99935382604599, '0': 0.0006462032906711102}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.9996199607849121, '0': 0.00037998449988663197}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.0015011822106316686, '0': 0.998498797416687}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.001304506789892912, '0': 0.9986955523490906}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.0001039106646203436, '0': 0.9998960494995117}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.0037654137704521418, '0': 0.9962345957756042}
Predicted Answer: 0, True Answer: 1
Probabilities for choices: {'1': 0.9912885427474976, '0': 0.008711383678019047}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.000656372809316963, '0': 0.9993435740470886}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.9997468590736389, '0': 0.0002531563804950565}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.00047285447362810373, '0': 0.9995272159576416}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.00029137005913071334, '0': 0.9997085928916931}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.9998623132705688, '0': 0.00013765462790615857}
Predicted Answer: 1, True Answer: 1
Evaluating: 100% | 872/872 [00:00<00:00, 2285.03it/s]
Evaluation completed. Accuracy: 0.9656 (842/872)
(base) root@e8e99ab71e21:~/Workspace/SF_CPT/SST-2#
```