Assignment3

2000017445 谢璐🄵

November 2023

1 Part 2

1.1 Bert Base

图 1: bert+base+restaurant

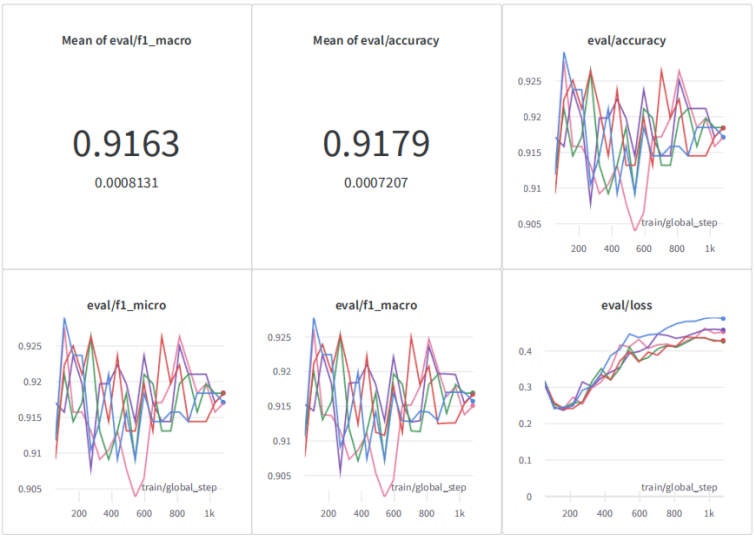

图 2: bert+base+agnews

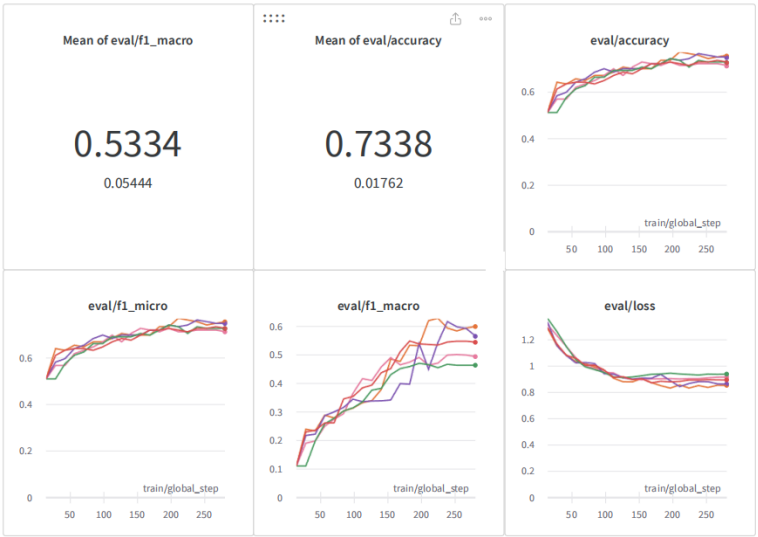

图 3: bert+base+acl

1.2 Roberta Base

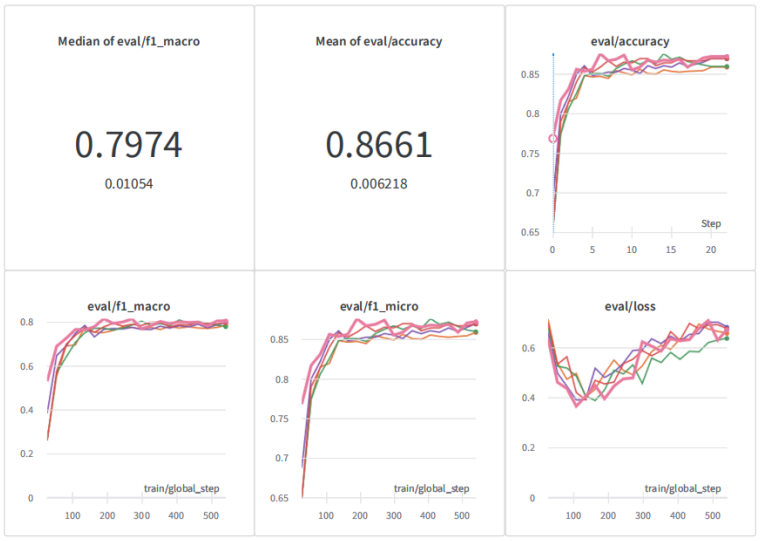

图 4: roberta+base+restaurant

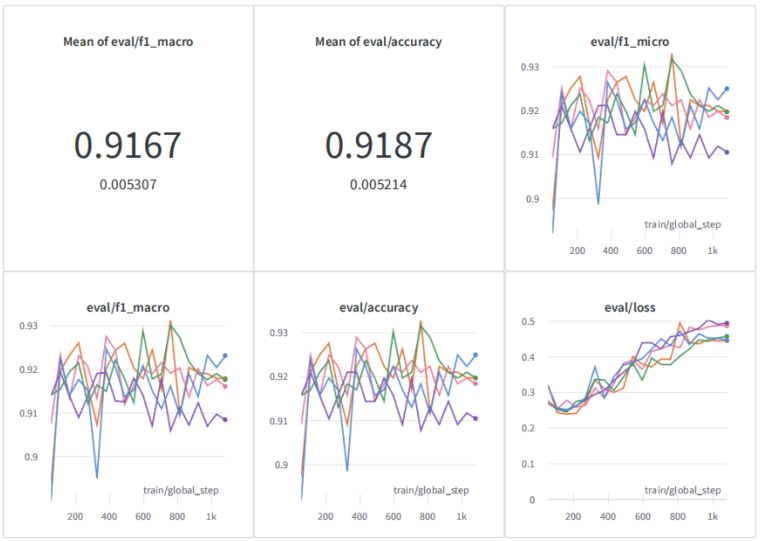

图 5: roberta+base+agnews

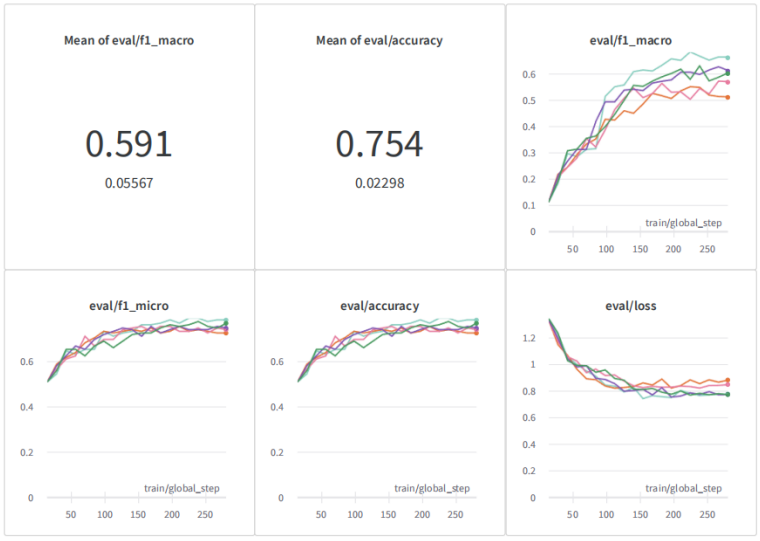

图 6: roberta+base+acl

1.3 Scibert

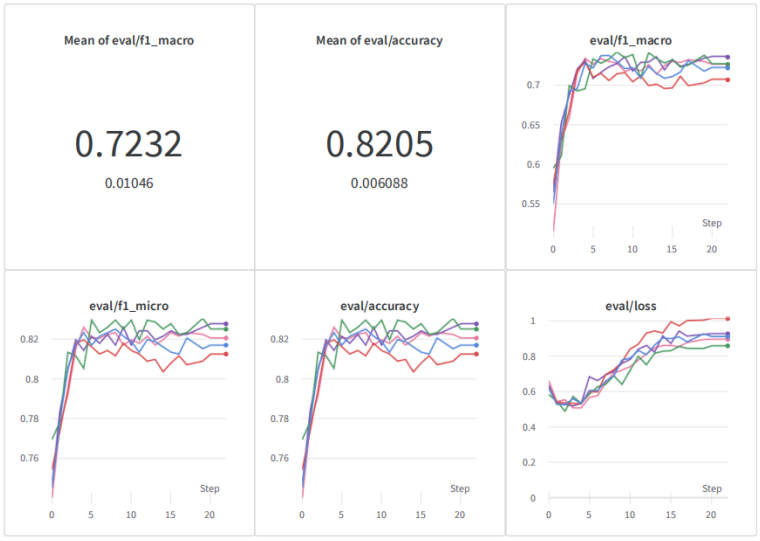

图 7: scibert+base+restaurant

图 8: scibert+acl

图 9: scibert+agnews

2 Part 3

2.1

sectionAdapter

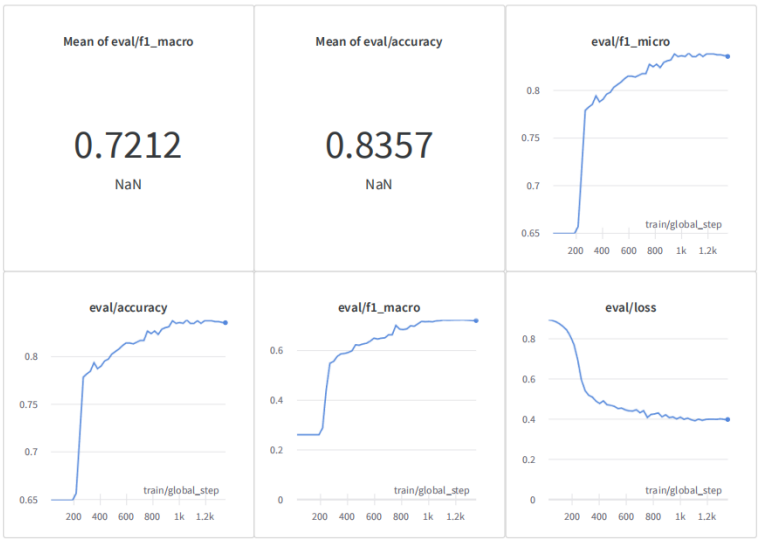

图 10: roberta+large+adapter+restaurant

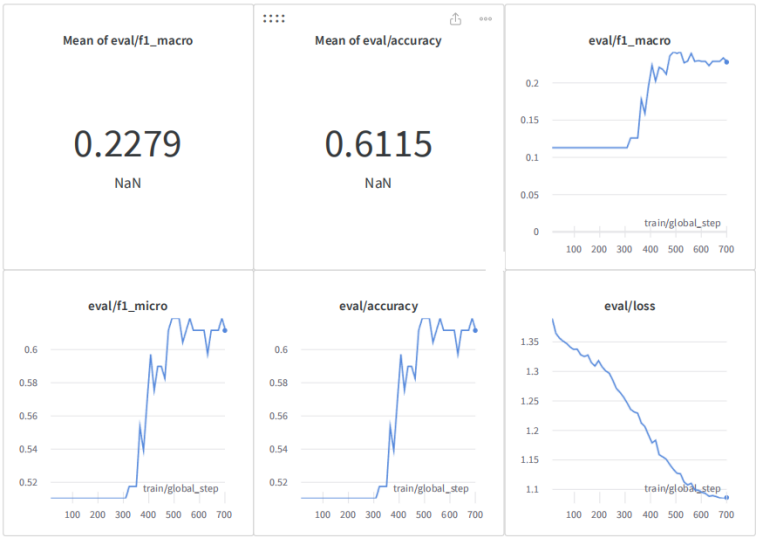

图 11: roberta+large+adapter+acl

图 12: roberta+large+adapter+agnews

2.2 Lora

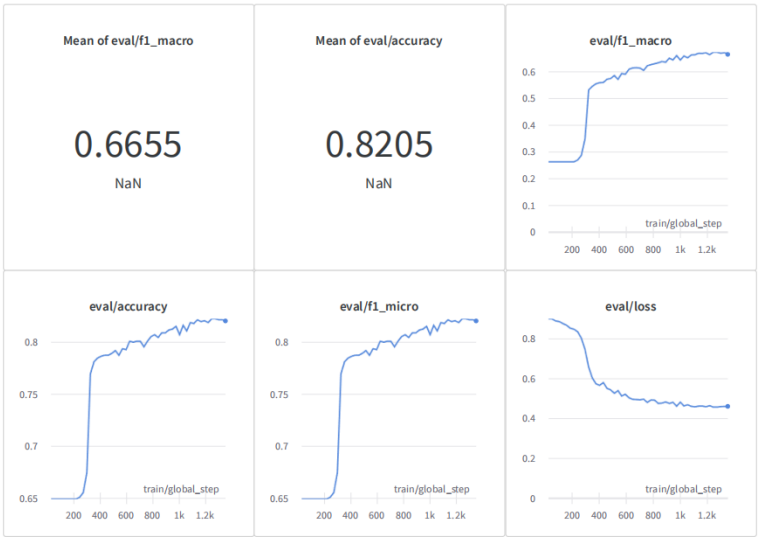

图 13: roberta+large+lora+restaurant

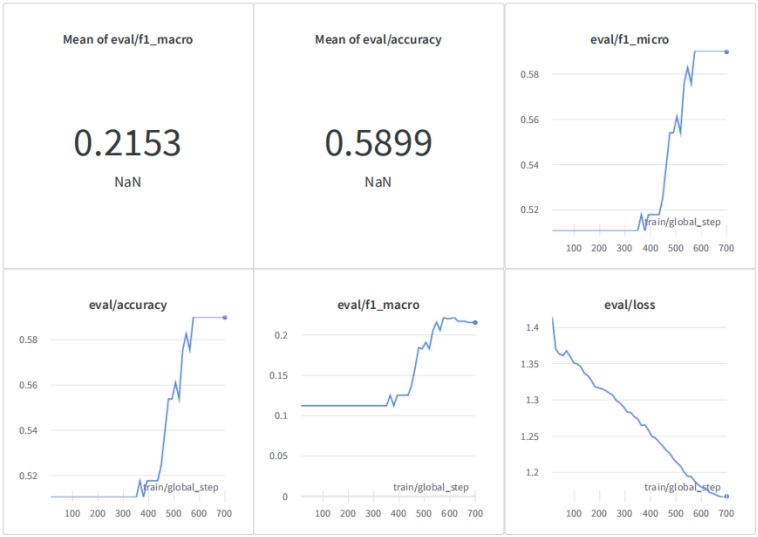

图 14: roberta+large+lora+acl

图 15: roberta+large+lora+agnews

2.3 内存占用分析

对于 OpenLLaMA-3B 模型，忽略其 embedding 层，其参数如下：

$hiddenstate := h = 3200$

$attentionhead := a = 32$

$layer := l = 26$

$batchsize := b = 64$

$p(parameter) \approx 12h^2 * l = 2,949,120,000 \approx 3 * 10^9$

假设每一个参数占 4byte，则占内存大小约为：

(1) parameters: $4 * p$

(2)gradient of every parameter: $4 * p$

(3)Optimizer states(Adam has two states per parameter): $4 * 2p$

(4)Intermediate activation: $4 * (34bh + 5b^2a)l$

$(1) + (2) + (3) = 48GB$

$(4) = 0.792GB$

如果采用 Lora 进行微调，对于每一层，假设 rank=16，Lora 对矩阵 q,k,v 作用：

$ratio = \frac{h*16*2*3}{12h^2} = 0.25\%$

(1),(2),(4) 占用内存不变，(3) 节省 99.75% 内存，故约节省 $23.94GB$ 内存!