Task3 Report

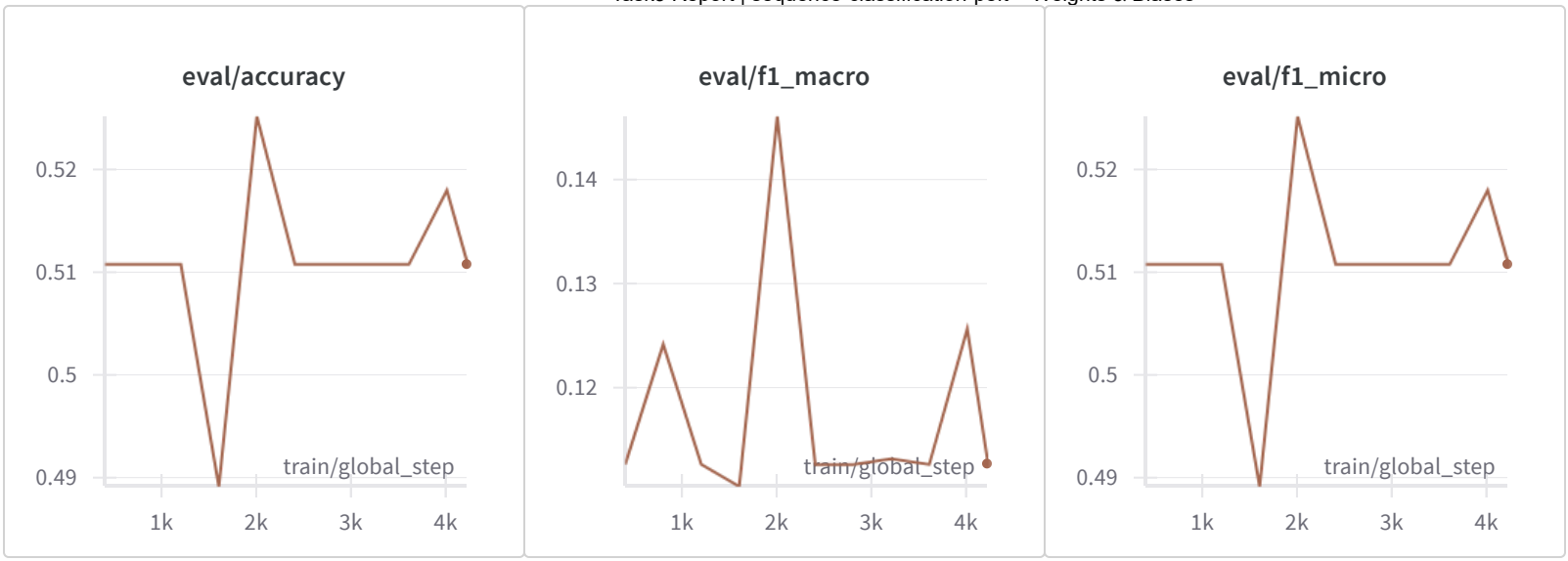
This is the report for task 3 in NLP-DLcourse

Jiahao Li

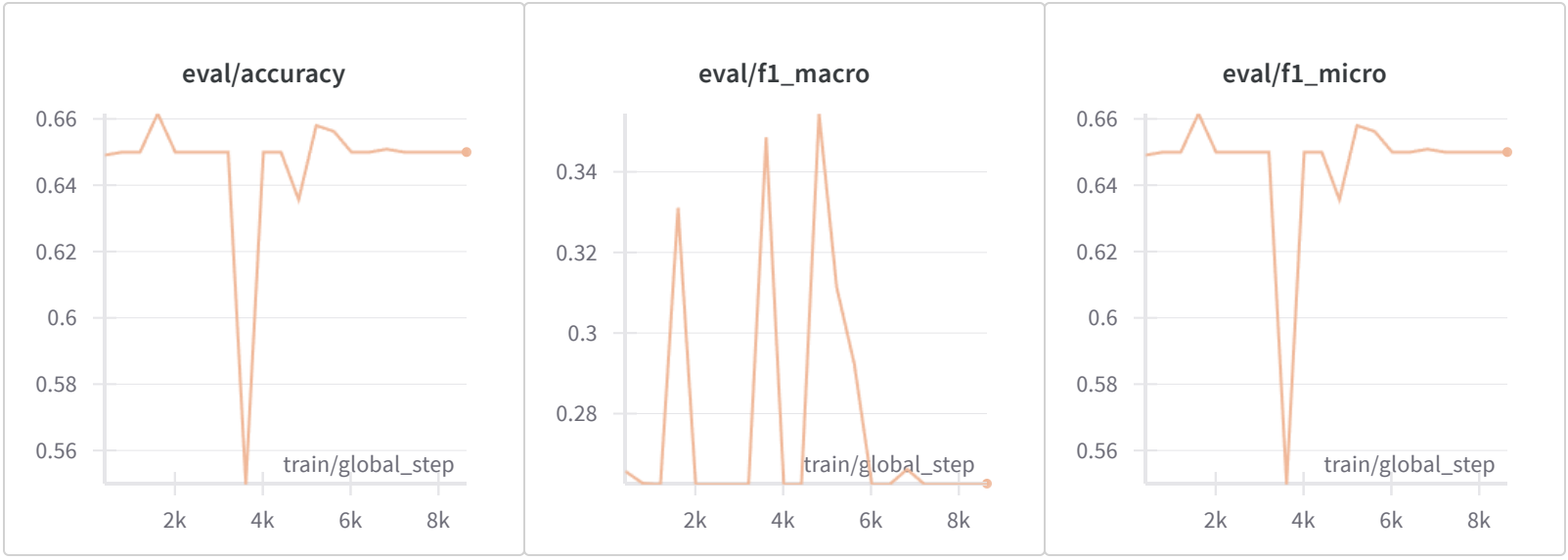
Created on October 19 | Last edited on November 4

Experiment

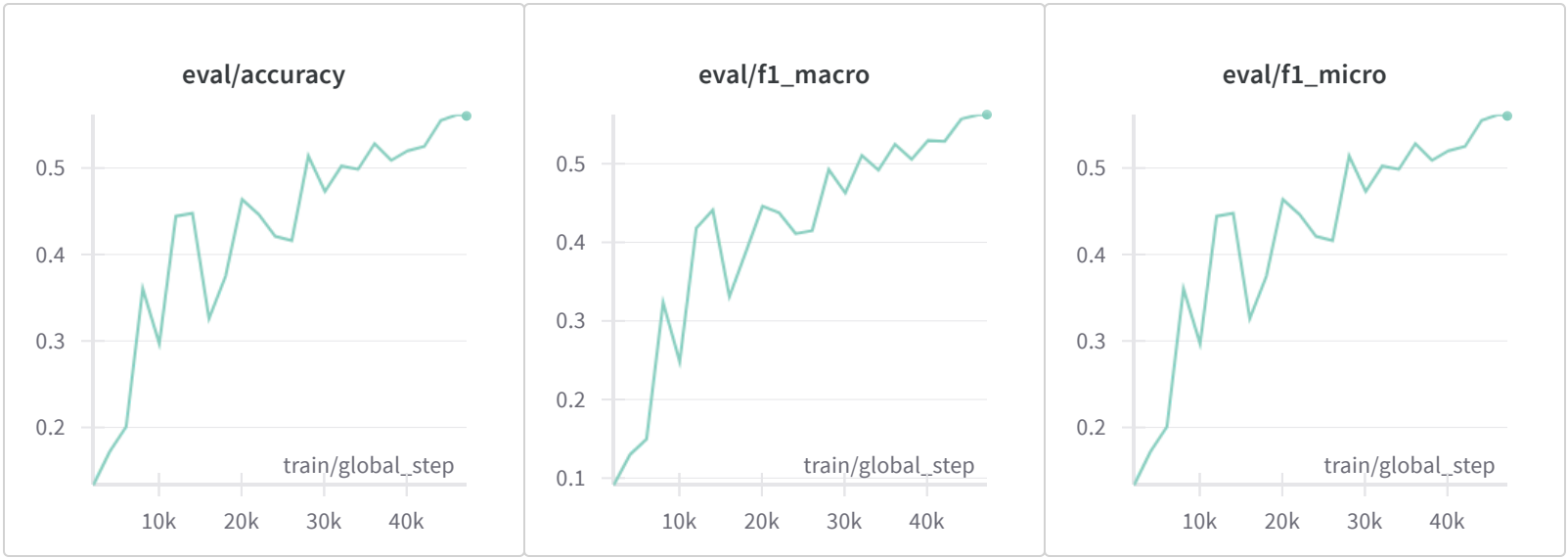
ACL dataset



Restaurant dataset



Agnews dataset



We can discover that adapters perform better on the agnews dataset while performing unsatisfyingly on the other two datasets. This is partly because the number of types of the agnews dataset is obviously more than that of the other two, so the adapters are able to change the prediction result.

In all the experiments, I set the hidden dimension of adapters to 256, and used the same hyperparameters as in task 2, except the epochs, which are set to 20.

Question

1. If I fine-tune a 3B model without PEFT, I will need approximately 30GB of GPU memory.
2. However, when I use PEFT(under my setting, an adapter whose hidden dimension is 256), I only need 6GB.

3. 80% GPU memory(24B) is saved.

Created with ❤️ on Weights & Biases.

https://wandb.ai/learnerljh-Peking%20University/sequence-classification-peft/reports/Task3-Report--Vmlldzo5NzgwNzc4