Date: 2023.12.10

Notes: The PyTorch template may be updated in the future, but the code in this example may not be affected by those changes.

SST2

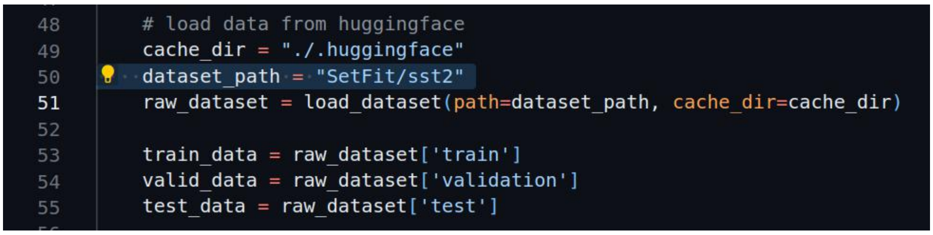
This is the record of how to tweak the pytorch template for this project, as well as what the training procedure looks like.

1. Tweak code

Remove the code unneeded, e.g., the code for cv.

1.1 About data

The SST2 dataset will be loaded from huggingface. Specify the dataset in **./main.py** and load the train, valid, and test splits.



```
    # load data from huggingface
    cache_dir = "./.huggingface"
    dataset_path = "SetFit/sst2"
    raw_dataset = load_dataset(path=dataset_path, cache_dir=cache_dir)

    train_data = raw_dataset['train']
    valid_data = raw_dataset['validation']
    test_data = raw_dataset['test']
```

The default function for preprocessing in **./preprocess.py** is fine, so I didn't touch it.

Also, the TextDataset class in **./dataset.py** could be used directly.

1.2 About model

The default MyModel in **./model.py** is designed for text classification task, I invoked it without any changes.

1.3 About training

By default, the template uses CrossEntropyLoss for criterion, AdamW for optimizer, CosineAnnealingWarmRestarts for lr scheduler, which seems appropriate. So I didn't touch these in **./main.py**, either.

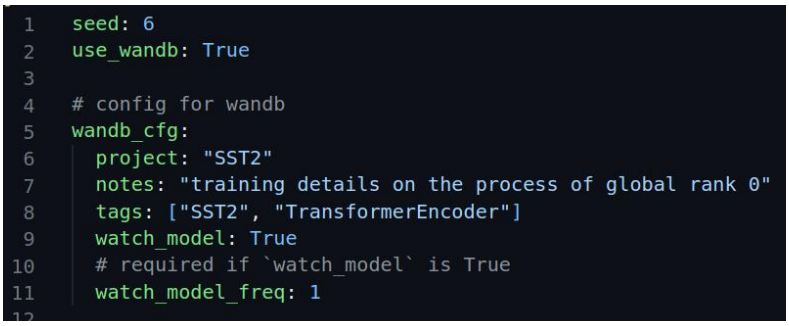
The Trainer in **./trainer.py** is ready-to-use, and it is recommended to use it directly without any alterations.

The template includes accuracy for test method, which just fit my demand in this simple project. So I kept it and didn't add more test methods.

1.4 About config

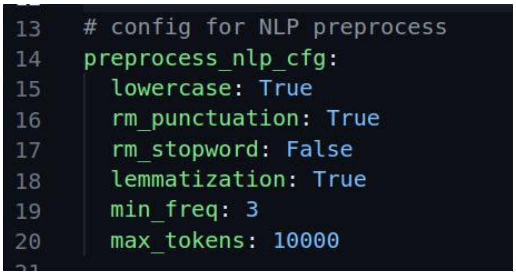
Tweak configurations in **./config.yaml**:

Use wandb to track experiment, set related config.



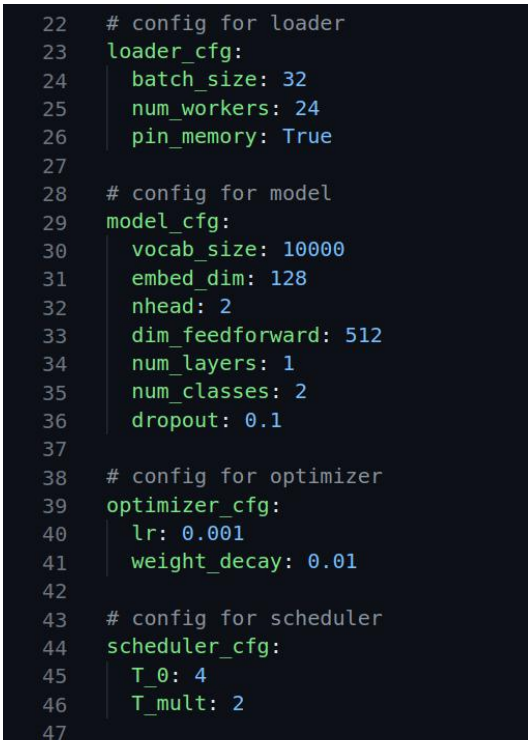
```yaml
seed: 6
use_wandb: True

# config for wandb
wandb_cfg:
  project: "SST2"
  notes: "training details on the process of global rank 0"
  tags: ["SST2", "TransformerEncoder"]
  watch_model: True
  # required if `watch_model` is True
  watch_model_freq: 1
```

Tweak config for preprocess, here the hyperparams are about tokenizer and vocabulary.



```yaml
# config for NLP preprocess
preprocess_nlp_cfg:
  lowercase: True
  rm_punctuation: True
  rm_stopword: False
  lemmatization: True
  min_freq: 3
  max_tokens: 10000
```

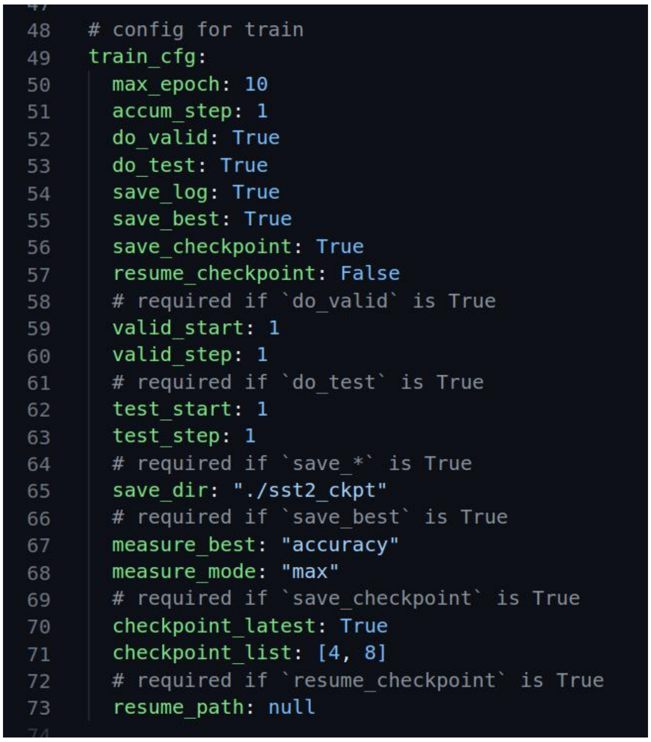
Tweak config for dataloader (e.g., batch_size), model (in this case, is a TransformerEncoder), optimizer (e.g., lr), and lr scheduler.



```yaml
# config for loader
loader_cfg:
  batch_size: 32
  num_workers: 24
  pin_memory: True

# config for model
model_cfg:
  vocab_size: 10000
  embed_dim: 128
  nhead: 2
  dim_feedforward: 512
  num_layers: 1
  num_classes: 2
  dropout: 0.1

# config for optimizer
optimizer_cfg:
  lr: 0.001
  weight_decay: 0.01

# config for scheduler
scheduler_cfg:
  T_0: 4
  T_mult: 2
```

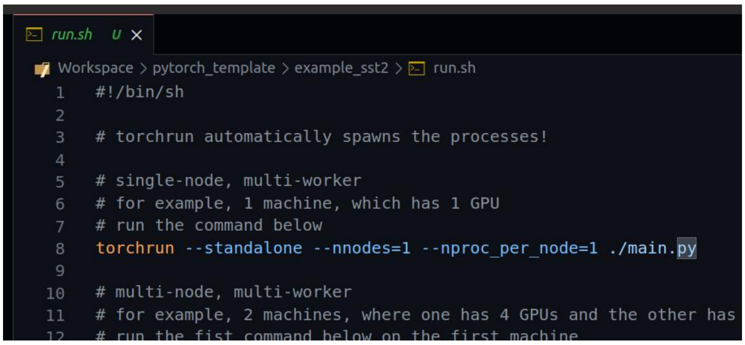
Tweak config for training. Here I made it to:

train up to 10 epochs;
not use gradient accumulation;
do validation and test accuracy;
save logs, best model, and checkpoints during training;
train from scratch rather than from checkpoint;
start validation at epoch 1 and at every 1 epoch;
start testing accuracy at epoch 1 and at every 1 epoch;
save logs and checkpoints to an existed directoty;
use accuracy to measure best model;
save latest checkpoint and checkpoints at specified epochs.



```
# config for train
train_cfg:
  max_epoch: 10
  accum_step: 1
  do_valid: True
  do_test: True
  save_log: True
  save_best: True
  save_checkpoint: True
  resume_checkpoint: False
  # required if `do_valid` is True
  valid_start: 1
  valid_step: 1
  # required if `do_test` is True
  test_start: 1
  test_step: 1
  # required if `save_*` is True
  save_dir: "./sst2_ckpt"
  # required if `save_best` is True
  measure_best: "accuracy"
  measure_mode: "max"
  # required if `save_checkpoint` is True
  checkpoint_latest: True
  checkpoint_list: [4, 8]
  # required if `resume_checkpoint` is True
  resume_path: null
```

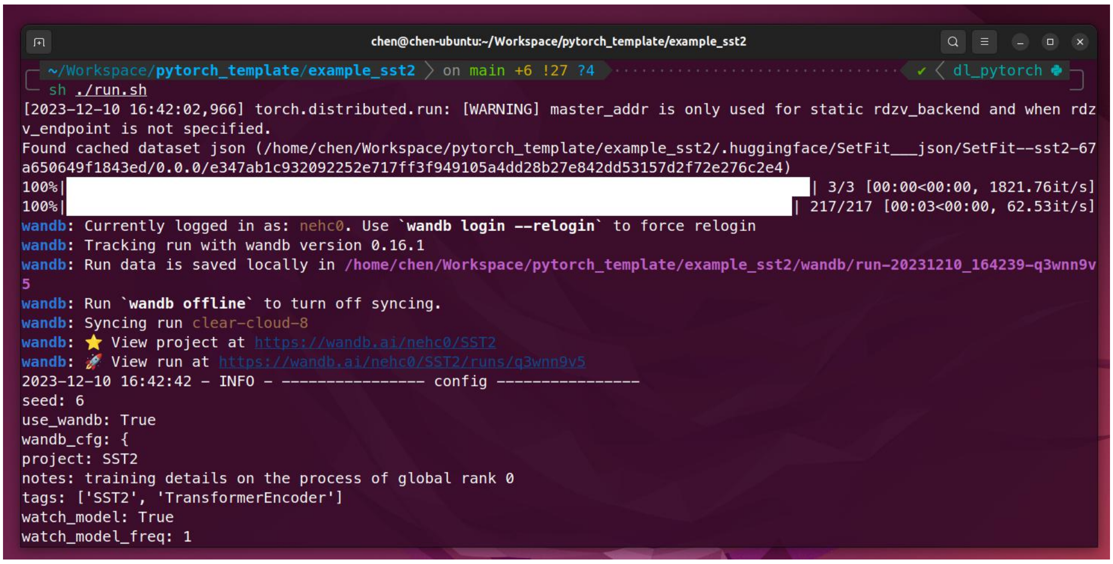
Finally, adjust **./run.sh** based on the machine architecture. I ran it on my laptop with single gpu.



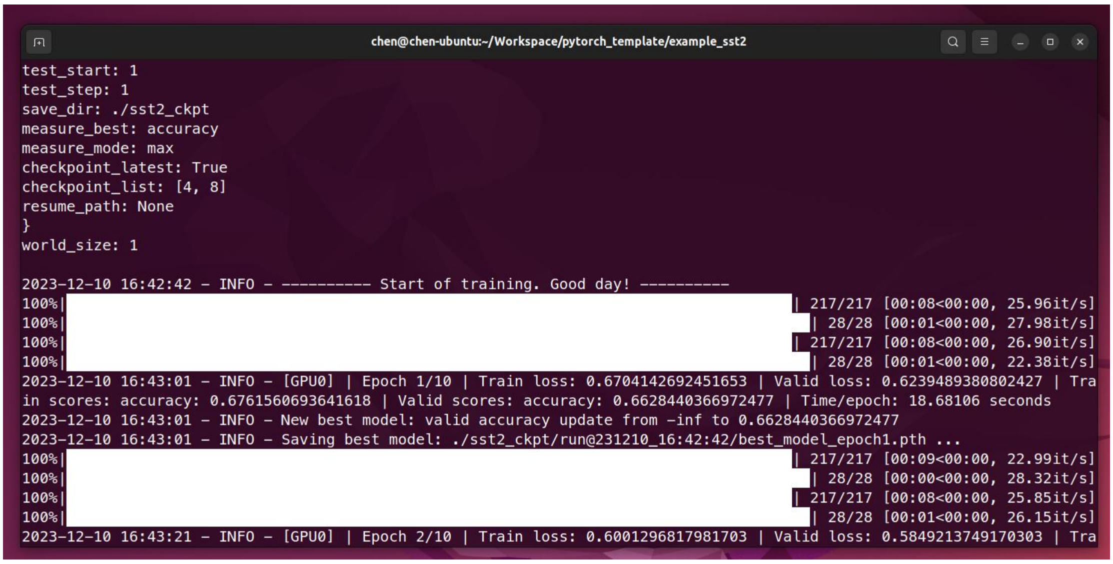
```
#!/bin/sh

# torchrun automatically spawns the processes!

# single-node, multi-worker
# for example, 1 machine, which has 1 GPU
# run the command below
torchrun --standalone --nnodes=1 --nproc_per_node=1 ./main.py

# multi-node, multi-worker
# for example, 2 machines, where one has 4 GPUs and the other has
# run the fist command below on the first machine
```

2. Training appearance

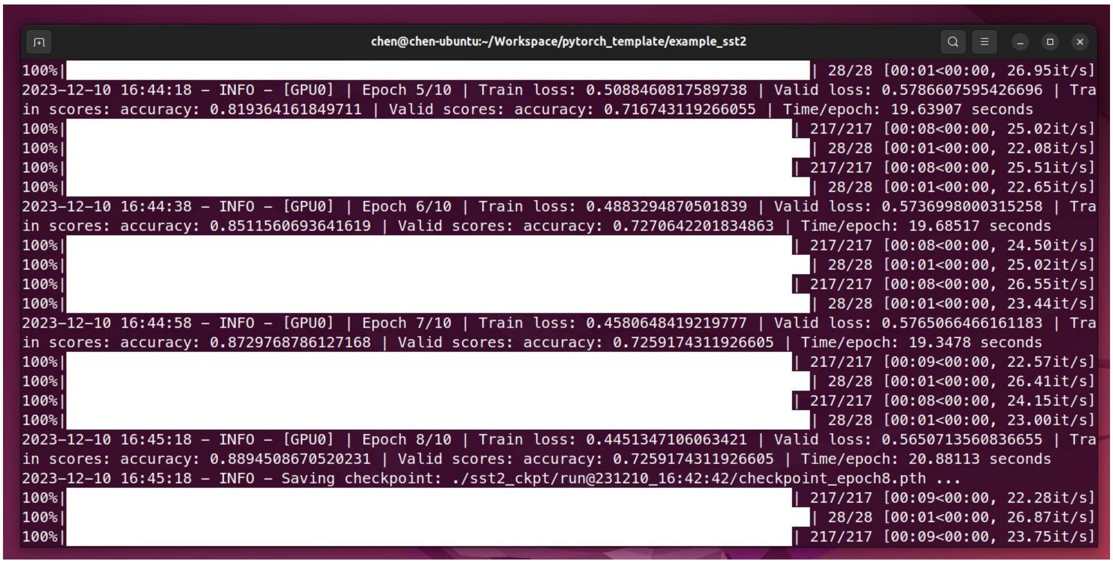
terminal:



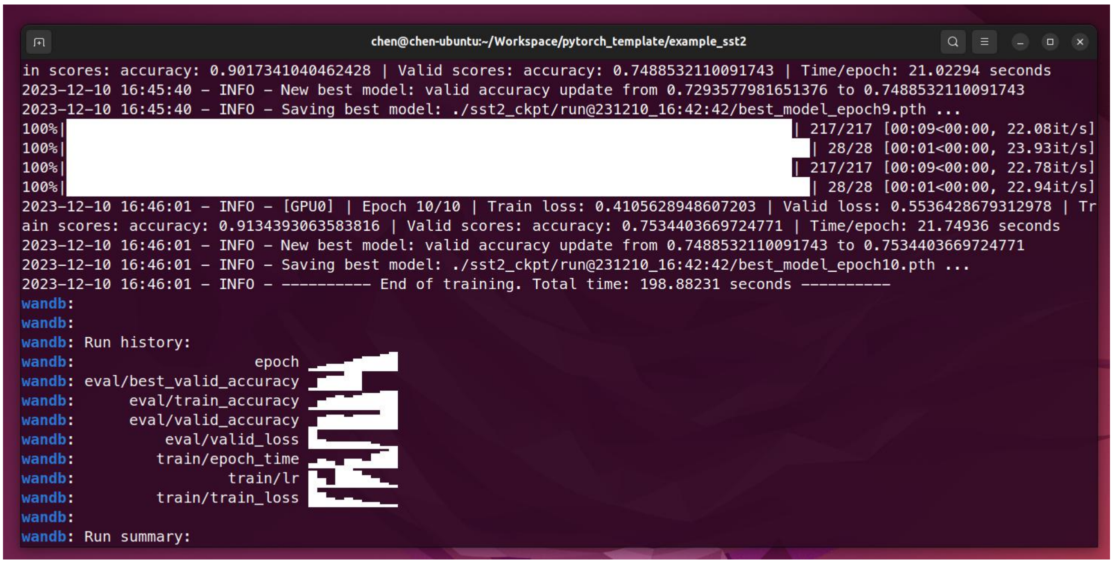
```
╭─ ~/Workspace/pytorch_template/example_sst2 ❯ on main +6 !27 ?4 ··························· ✔ ❮ dl_pytorch ─╮
╰─ sh ./run.sh
[2023-12-10 16:42:02,966] torch.distributed.run: [WARNING] master_addr is only used for static rdzv_backend and when rdz
v_endpoint is not specified.
Found cached dataset json (/home/chen/Workspace/pytorch_template/example_sst2/.huggingface/SetFit___json/SetFit--sst2-67
a650649f1843ed/0.0.0/e347ab1c932092252e717ff3f949105a4dd28b27e842dd53157d2f72e276c2e4)
100%|██████████████████████████████████████████████████████████████████████████████████| 3/3 [00:00<00:00, 1821.76it/s]
100%|████████████████████████████████████████████████████████████████████████████████| 217/217 [00:03<00:00, 62.53it/s]
wandb: Currently logged in as: nehc0. Use `wandb login --relogin` to force relogin
wandb: Tracking run with wandb version 0.16.1
wandb: Run data is saved locally in /home/chen/Workspace/pytorch_template/example_sst2/wandb/run-20231210_164239-q3wnn9v
5
wandb: Run `wandb offline` to turn off syncing.
wandb: Syncing run clear-cloud-8
wandb: ⭐️ View project at https://wandb.ai/nehc0/SST2
wandb: 🚀 View run at https://wandb.ai/nehc0/SST2/runs/q3wnn9v5
2023-12-10 16:42:42 - INFO - ---------------- config ----------------
seed: 6
use_wandb: True
wandb_cfg: {
project: SST2
notes: training details on the process of global rank 0
tags: ['SST2', 'TransformerEncoder']
watch_model: True
watch_model_freq: 1
```



```
test_start: 1
test_step: 1
save_dir: ./sst2_ckpt
measure_best: accuracy
measure_mode: max
checkpoint_latest: True
checkpoint_list: [4, 8]
resume_path: None
}
world_size: 1

2023-12-10 16:42:42 - INFO - ---------- Start of training. Good day! ----------
100%|████████████████████████████████████████████████████████████████████████████████| 217/217 [00:08<00:00, 25.96it/s]
100%|██████████████████████████████████████████████████████████████████████████████████| 28/28 [00:01<00:00, 27.98it/s]
100%|████████████████████████████████████████████████████████████████████████████████| 217/217 [00:08<00:00, 26.90it/s]
100%|██████████████████████████████████████████████████████████████████████████████████| 28/28 [00:01<00:00, 22.38it/s]
2023-12-10 16:43:01 - INFO - [GPU0] | Epoch 1/10 | Train loss: 0.6704142692451653 | Valid loss: 0.6239489380802427 | Tra
in scores: accuracy: 0.6761560693641618 | Valid scores: accuracy: 0.6628440366972477 | Time/epoch: 18.68106 seconds
2023-12-10 16:43:01 - INFO - New best model: valid accuracy update from -inf to 0.6628440366972477
2023-12-10 16:43:01 - INFO - Saving best model: ./sst2_ckpt/run@231210_16:42:42/best_model_epoch1.pth ...
100%|████████████████████████████████████████████████████████████████████████████████| 217/217 [00:09<00:00, 22.99it/s]
100%|██████████████████████████████████████████████████████████████████████████████████| 28/28 [00:00<00:00, 28.32it/s]
100%|████████████████████████████████████████████████████████████████████████████████| 217/217 [00:08<00:00, 25.85it/s]
100%|██████████████████████████████████████████████████████████████████████████████████| 28/28 [00:01<00:00, 26.15it/s]
2023-12-10 16:43:21 - INFO - [GPU0] | Epoch 2/10 | Train loss: 0.6001296817981703 | Valid loss: 0.5849213749170303 | Tra
```

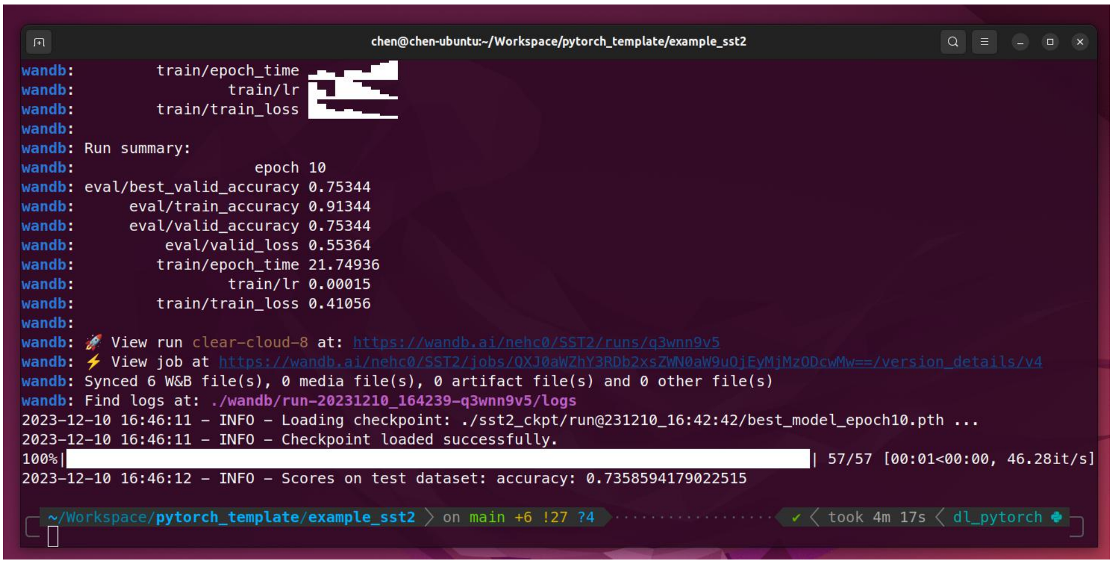


```
100%|██████████████████████████████████████████████████████████████████████████████████| 28/28 [00:01<00:00, 26.95it/s]
2023-12-10 16:44:18 - INFO - [GPU0] | Epoch 5/10 | Train loss: 0.5088460817589738 | Valid loss: 0.5786607595426696 | Tra
in scores: accuracy: 0.819364161849711 | Valid scores: accuracy: 0.716743119266055 | Time/epoch: 19.63907 seconds
100%|████████████████████████████████████████████████████████████████████████████████| 217/217 [00:08<00:00, 25.02it/s]
100%|██████████████████████████████████████████████████████████████████████████████████| 28/28 [00:01<00:00, 22.08it/s]
100%|████████████████████████████████████████████████████████████████████████████████| 217/217 [00:08<00:00, 25.51it/s]
100%|██████████████████████████████████████████████████████████████████████████████████| 28/28 [00:01<00:00, 22.65it/s]
2023-12-10 16:44:38 - INFO - [GPU0] | Epoch 6/10 | Train loss: 0.4883294870501839 | Valid loss: 0.5736998000315258 | Tra
in scores: accuracy: 0.8511560693641619 | Valid scores: accuracy: 0.7270642201834863 | Time/epoch: 19.68517 seconds
100%|████████████████████████████████████████████████████████████████████████████████| 217/217 [00:08<00:00, 24.50it/s]
100%|██████████████████████████████████████████████████████████████████████████████████| 28/28 [00:01<00:00, 25.02it/s]
100%|████████████████████████████████████████████████████████████████████████████████| 217/217 [00:08<00:00, 26.55it/s]
100%|██████████████████████████████████████████████████████████████████████████████████| 28/28 [00:01<00:00, 23.44it/s]
2023-12-10 16:44:58 - INFO - [GPU0] | Epoch 7/10 | Train loss: 0.4580648419219777 | Valid loss: 0.5765066466161183 | Tra
in scores: accuracy: 0.8729768786127168 | Valid scores: accuracy: 0.7259174311926605 | Time/epoch: 19.3478 seconds
100%|████████████████████████████████████████████████████████████████████████████████| 217/217 [00:09<00:00, 22.57it/s]
100%|██████████████████████████████████████████████████████████████████████████████████| 28/28 [00:01<00:00, 26.41it/s]
100%|████████████████████████████████████████████████████████████████████████████████| 217/217 [00:08<00:00, 24.15it/s]
100%|██████████████████████████████████████████████████████████████████████████████████| 28/28 [00:01<00:00, 23.00it/s]
2023-12-10 16:45:18 - INFO - [GPU0] | Epoch 8/10 | Train loss: 0.4451347106063421 | Valid loss: 0.5650713560836655 | Tra
in scores: accuracy: 0.8894508670520231 | Valid scores: accuracy: 0.7259174311926605 | Time/epoch: 20.88113 seconds
2023-12-10 16:45:18 - INFO - Saving checkpoint: ./sst2_ckpt/run@231210_16:42:42/checkpoint_epoch8.pth ...
100%|████████████████████████████████████████████████████████████████████████████████| 217/217 [00:09<00:00, 22.28it/s]
100%|██████████████████████████████████████████████████████████████████████████████████| 28/28 [00:01<00:00, 26.87it/s]
100%|████████████████████████████████████████████████████████████████████████████████| 217/217 [00:09<00:00, 23.75it/s]
```



chen@chen-ubuntu:~/Workspace/pytorch_template/example_sst2

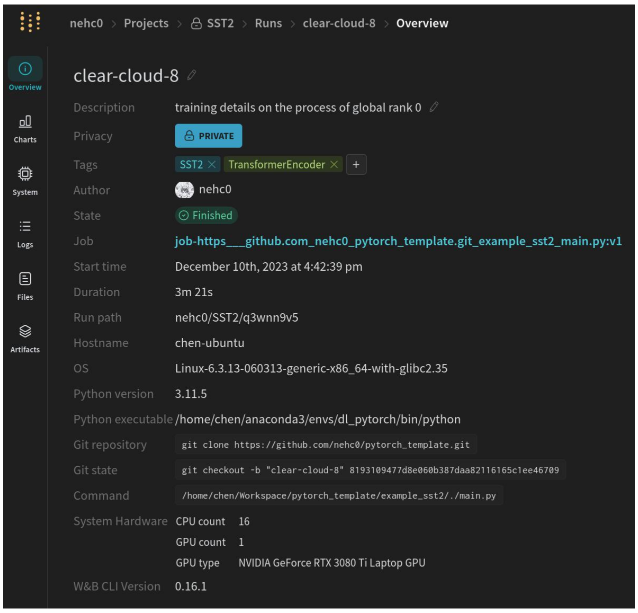
```
in scores: accuracy: 0.9017341040462428 | Valid scores: accuracy: 0.7488532110091743 | Time/epoch: 21.02294 seconds
2023-12-10 16:45:40 - INFO - New best model: valid accuracy update from 0.7293577981651376 to 0.7488532110091743
2023-12-10 16:45:40 - INFO - Saving best model: ./sst2_ckpt/run@231210_16:42:42/best_model_epoch9.pth ...
100%|██████████████████████████████████████████████████████████████████████| 217/217 [00:09<00:00, 22.08it/s]
100%|████████████████████████████████████████████████████████████████████████| 28/28 [00:01<00:00, 23.93it/s]
100%|██████████████████████████████████████████████████████████████████████| 217/217 [00:09<00:00, 22.78it/s]
100%|████████████████████████████████████████████████████████████████████████| 28/28 [00:01<00:00, 22.94it/s]
2023-12-10 16:46:01 - INFO - [GPU0] | Epoch 10/10 | Train loss: 0.4105628948607203 | Valid loss: 0.5536428679312978 | Tr
ain scores: accuracy: 0.9134393063583816 | Valid scores: accuracy: 0.7534403669724771 | Time/epoch: 21.74936 seconds
2023-12-10 16:46:01 - INFO - New best model: valid accuracy update from 0.7488532110091743 to 0.7534403669724771
2023-12-10 16:46:01 - INFO - Saving best model: ./sst2_ckpt/run@231210_16:42:42/best_model_epoch10.pth ...
2023-12-10 16:46:01 - INFO - ---------- End of training. Total time: 198.88231 seconds ----------
wandb:
wandb:
wandb: Run history:
wandb:                    epoch
wandb: eval/best_valid_accuracy
wandb:      eval/train_accuracy
wandb:      eval/valid_accuracy
wandb:          eval/valid_loss
wandb:         train/epoch_time
wandb:                 train/lr
wandb:         train/train_loss
wandb:
wandb: Run summary:
```



chen@chen-ubuntu:~/Workspace/pytorch_template/example_sst2

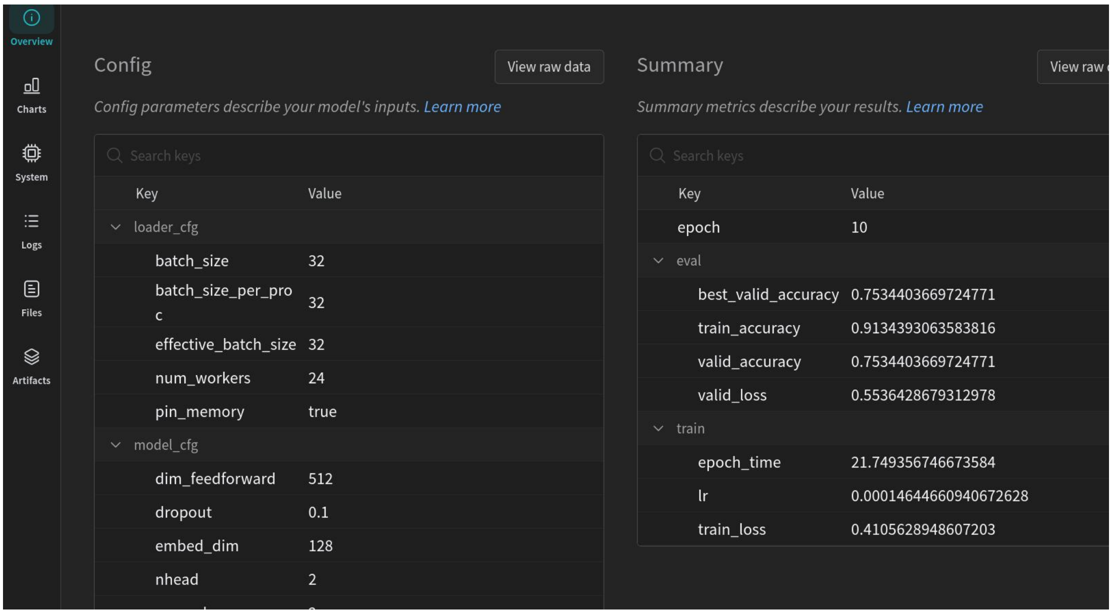
```
wandb:         train/epoch_time
wandb:                 train/lr
wandb:         train/train_loss
wandb:
wandb: Run summary:
wandb:                    epoch 10
wandb: eval/best_valid_accuracy 0.75344
wandb:      eval/train_accuracy 0.91344
wandb:      eval/valid_accuracy 0.75344
wandb:          eval/valid_loss 0.55364
wandb:         train/epoch_time 21.74936
wandb:                 train/lr 0.00015
wandb:         train/train_loss 0.41056
wandb:
wandb: 🚀 View run clear-cloud-8 at: https://wandb.ai/nehc0/SST2/runs/q3wnn9v5
wandb: ⚡ View job at https://wandb.ai/nehc0/SST2/jobs/QXJ0aWZhY3RDb2xsZWN0aW9uOjEyMjMzODcwMw==/version_details/v4
wandb: Synced 6 W&B file(s), 0 media file(s), 0 artifact file(s) and 0 other file(s)
wandb: Find logs at: ./wandb/run-20231210_164239-q3wnn9v5/logs
2023-12-10 16:46:11 - INFO - Loading checkpoint: ./sst2_ckpt/run@231210_16:42:42/best_model_epoch10.pth ...
2023-12-10 16:46:11 - INFO - Checkpoint loaded successfully.
100%|████████████████████████████████████████████████████████████████████████| 57/57 [00:01<00:00, 46.28it/s]
2023-12-10 16:46:12 - INFO - Scores on test dataset: accuracy: 0.7358594179022515

 ~/Workspace/pytorch_template/example_sst2 〉 on main +6 !27 ?4 ·················· ✔ 〈 took 4m 17s 〈 dl_pytorch
```

wandb:



clear-cloud-8

Description	training details on the process of global rank 0
Privacy	PRIVATE
Tags	SST2 × TransformerEncoder × +
Author	nehc0
State	Finished
Job	job-https___github.com_nehc0_pytorch_template.git_example_sst2_main.py:v1
Start time	December 10th, 2023 at 4:42:39 pm
Duration	3m 21s
Run path	nehc0/SST2/q3wnn9v5
Hostname	chen-ubuntu
OS	Linux-6.3.13-060313-generic-x86_64-with-glibc2.35
Python version	3.11.5
Python executable	/home/chen/anaconda3/envs/dl_pytorch/bin/python
Git repository	git clone https://github.com/nehc0/pytorch_template.git
Git state	git checkout -b "clear-cloud-8" 8193109477d8e060b387daa82116165c1ee46709
Command	/home/chen/Workspace/pytorch_template/example_sst2/./main.py
System Hardware	CPU count 16 GPU count 1 GPU type NVIDIA GeForce RTX 3080 Ti Laptop GPU
W&B CLI Version	0.16.1



Config

Config parameters describe your model's inputs. Learn more

Search keys

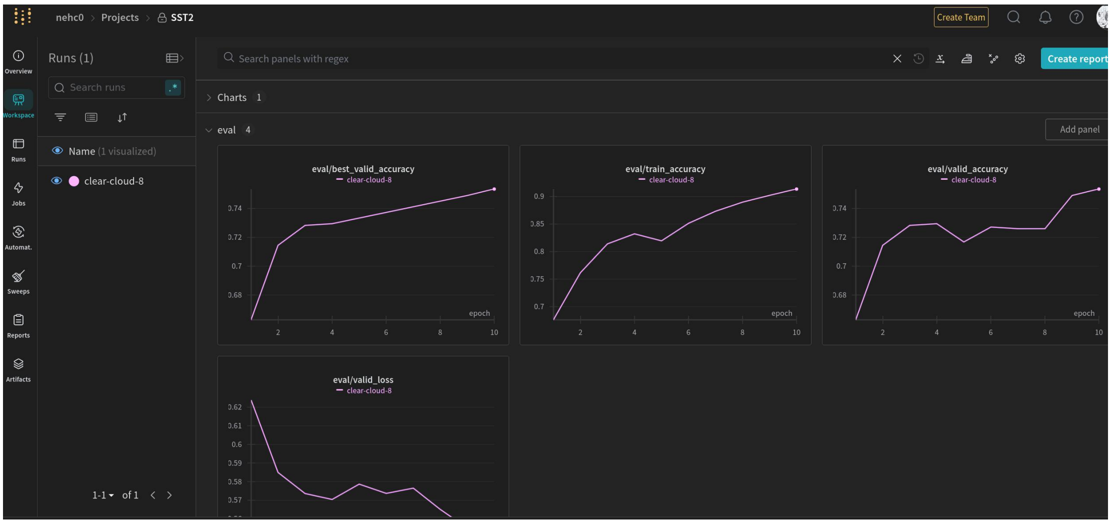
Key	Value
loader_cfg	
batch_size	32
batch_size_per_proc	32
effective_batch_size	32
num_workers	24
pin_memory	true
model_cfg	
dim_feedforward	512
dropout	0.1
embed_dim	128
nhead	2

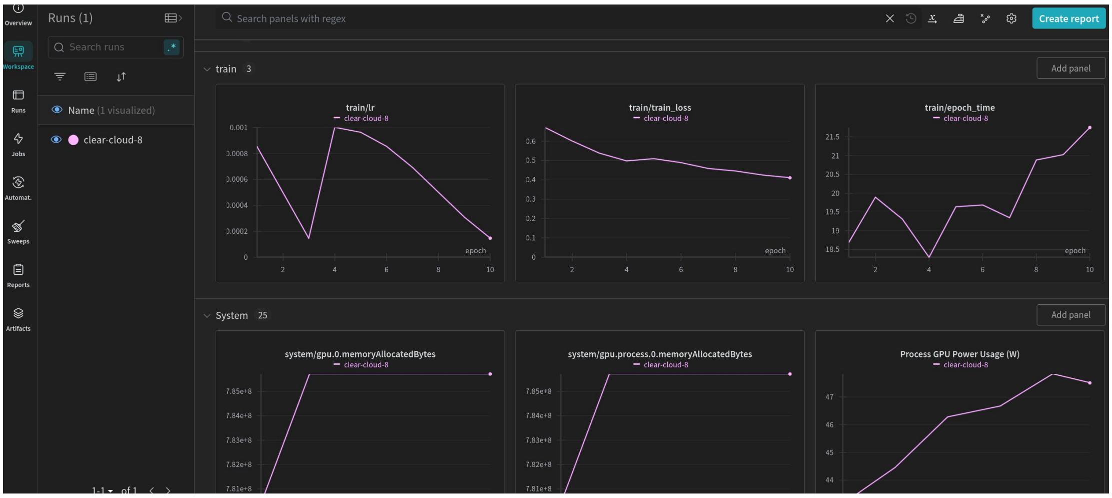
Summary

Summary metrics describe your results. Learn more

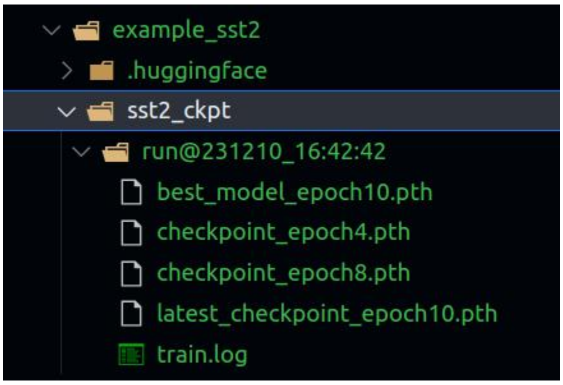
Search keys

Key	Value
epoch	10
eval	
best_valid_accuracy	0.7534403669724771
train_accuracy	0.9134393063583816
valid_accuracy	0.7534403669724771
valid_loss	0.5536428679312978
train	
epoch_time	21.749356746673584
lr	0.00014644660940672628
train_loss	0.4105628948607203

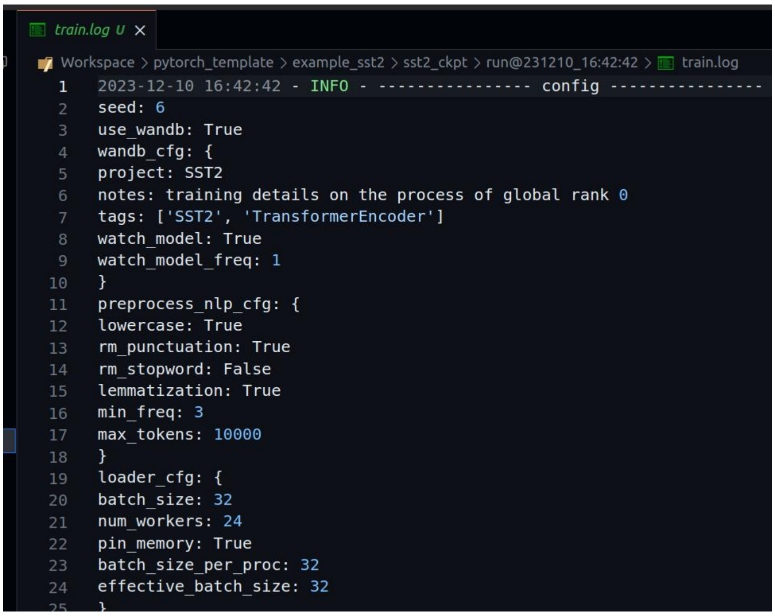




saved logs and checkpoints:

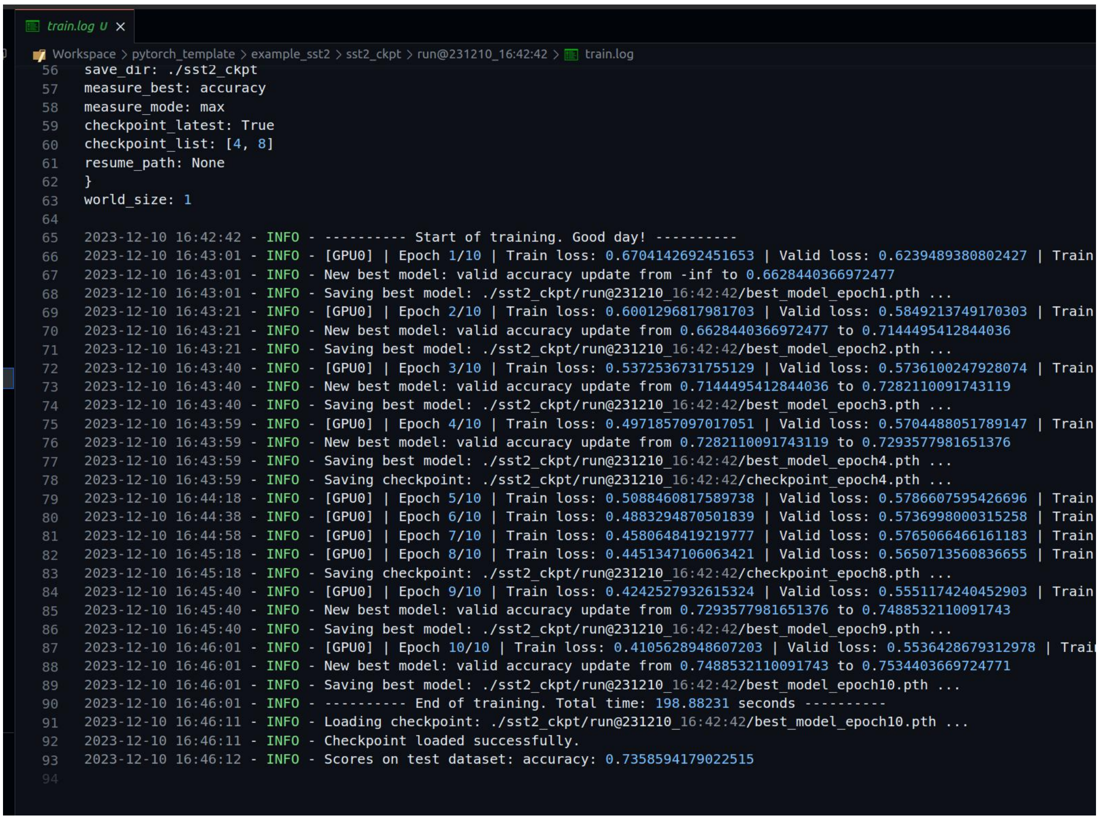


```
example_sst2
  .huggingface
  sst2_ckpt
    run@231210_16:42:42
      best_model_epoch10.pth
      checkpoint_epoch4.pth
      checkpoint_epoch8.pth
      latest_checkpoint_epoch10.pth
      train.log
```



train.log U ×
Workspace > pytorch_template > example_sst2 > sst2_ckpt > run@231210_16:42:42 > train.log

```
2023-12-10 16:42:42 - INFO - ---------------- config ----------------
seed: 6
use_wandb: True
wandb_cfg: {
project: SST2
notes: training details on the process of global rank 0
tags: ['SST2', 'TransformerEncoder']
watch_model: True
watch_model_freq: 1
}
preprocess_nlp_cfg: {
lowercase: True
rm_punctuation: True
rm_stopword: False
lemmatization: True
min_freq: 3
max_tokens: 10000
}
loader_cfg: {
batch_size: 32
num_workers: 24
pin_memory: True
batch_size_per_proc: 32
effective_batch_size: 32
}
```



train.log U ×
Workspace > pytorch_template > example_sst2 > sst2_ckpt > run@231210_16:42:42 > train.log

```
save_dir: ./sst2_ckpt
measure_best: accuracy
measure_mode: max
checkpoint_latest: True
checkpoint_list: [4, 8]
resume_path: None
}
world_size: 1

2023-12-10 16:42:42 - INFO - ---------- Start of training. Good day! ----------
2023-12-10 16:43:01 - INFO - [GPU0] | Epoch 1/10 | Train loss: 0.6704142692451653 | Valid loss: 0.6239489380802427 | Train
2023-12-10 16:43:01 - INFO - New best model: valid accuracy update from -inf to 0.6628440366972477
2023-12-10 16:43:01 - INFO - Saving best model: ./sst2_ckpt/run@231210_16:42:42/best_model_epoch1.pth ...
2023-12-10 16:43:21 - INFO - [GPU0] | Epoch 2/10 | Train loss: 0.6001296817981703 | Valid loss: 0.5849213749170303 | Train
2023-12-10 16:43:21 - INFO - New best model: valid accuracy update from 0.6628440366972477 to 0.7144495412844036
2023-12-10 16:43:21 - INFO - Saving best model: ./sst2_ckpt/run@231210_16:42:42/best_model_epoch2.pth ...
2023-12-10 16:43:40 - INFO - [GPU0] | Epoch 3/10 | Train loss: 0.5372536731755129 | Valid loss: 0.5736100247928074 | Train
2023-12-10 16:43:40 - INFO - New best model: valid accuracy update from 0.7144495412844036 to 0.7282110091743119
2023-12-10 16:43:40 - INFO - Saving best model: ./sst2_ckpt/run@231210_16:42:42/best_model_epoch3.pth ...
2023-12-10 16:43:59 - INFO - [GPU0] | Epoch 4/10 | Train loss: 0.4971857097017051 | Valid loss: 0.5704488051789147 | Train
2023-12-10 16:43:59 - INFO - New best model: valid accuracy update from 0.7282110091743119 to 0.7293577981651376
2023-12-10 16:43:59 - INFO - Saving best model: ./sst2_ckpt/run@231210_16:42:42/best_model_epoch4.pth ...
2023-12-10 16:43:59 - INFO - Saving checkpoint: ./sst2_ckpt/run@231210_16:42:42/checkpoint_epoch4.pth ...
2023-12-10 16:44:18 - INFO - [GPU0] | Epoch 5/10 | Train loss: 0.5088460817589738 | Valid loss: 0.5786607595426696 | Train
2023-12-10 16:44:38 - INFO - [GPU0] | Epoch 6/10 | Train loss: 0.4883294870501839 | Valid loss: 0.5736998000315258 | Train
2023-12-10 16:44:58 - INFO - [GPU0] | Epoch 7/10 | Train loss: 0.4580648419219777 | Valid loss: 0.5765066466161183 | Train
2023-12-10 16:45:18 - INFO - [GPU0] | Epoch 8/10 | Train loss: 0.4451347106063421 | Valid loss: 0.5650713560836655 | Train
2023-12-10 16:45:18 - INFO - Saving checkpoint: ./sst2_ckpt/run@231210_16:42:42/checkpoint_epoch8.pth ...
2023-12-10 16:45:40 - INFO - [GPU0] | Epoch 9/10 | Train loss: 0.4242527932615324 | Valid loss: 0.5551174240452903 | Train
2023-12-10 16:45:40 - INFO - New best model: valid accuracy update from 0.7293577981651376 to 0.7488532110091743
2023-12-10 16:45:40 - INFO - Saving best model: ./sst2_ckpt/run@231210_16:42:42/best_model_epoch9.pth ...
2023-12-10 16:46:01 - INFO - [GPU0] | Epoch 10/10 | Train loss: 0.4105628948607203 | Valid loss: 0.5536428679312978 | Trai
2023-12-10 16:46:01 - INFO - New best model: valid accuracy update from 0.7488532110091743 to 0.7534403669724771
2023-12-10 16:46:01 - INFO - Saving best model: ./sst2_ckpt/run@231210_16:42:42/best_model_epoch10.pth ...
2023-12-10 16:46:01 - INFO - ---------- End of training. Total time: 198.88231 seconds ----------
2023-12-10 16:46:11 - INFO - Loading checkpoint: ./sst2_ckpt/run@231210_16:42:42/best_model_epoch10.pth ...
2023-12-10 16:46:11 - INFO - Checkpoint loaded successfully.
2023-12-10 16:46:12 - INFO - Scores on test dataset: accuracy: 0.7358594179022515
```