

total_frozen_params=210236928
trainable parameters: 28366848
/tmp/ipython-input-32-347286690.py:16: FutureWarning: `torch.cuda.amp.GradScaler(args...)` is deprecated. Please use `torch.amp.GradScaler('cuda', args...)` instead.
self.scaler = GradScaler()
/tmp/ipython-input-32-347286690.py:38: UserWarning: Argument(s) 'always_apply' are not valid for transform Normalize
A.Normalize(mean=[0.5,0.5,0.5],std=[0.5,0.5,0.5],always_apply=True),
DistillationTrainer initialized. Ready to create a distilled dataset.
Loading pre-trained model for distillation...
Creating distilled dataset with top 5.0% hardest examples...

Calculating losses for distillation: 100% 782/782 [06:37<00:00, 1.95it/s]

Original dataset size: 100000
Distilled dataset size: 5000

--- Training on DISTILLED dataset from scratch ---
total_frozen_params=210236928
trainable parameters: 28366848
/tmp/ipython-input-32-347286690.py:16: FutureWarning: `torch.cuda.amp.GradScaler(args...)` is deprecated. Please use `torch.amp.GradScaler('cuda', args...)` instead.
self.scaler = GradScaler()
/tmp/ipython-input-32-347286690.py:38: UserWarning: Argument(s) 'always_apply' are not valid for transform Normalize
A.Normalize(mean=[0.5,0.5,0.5],std=[0.5,0.5,0.5],always_apply=True),
validating: 100% 10/10 [12:51<00:00, 79.74s/it]

train loss: 8.738: 100% 40/40 [00:25<00:00, 2.50it/s]
/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
with autocast():
valid loss: 8.591: 100% 79/79 [00:30<00:00, 3.87it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
with autocast():
train_loss train_perplexity val_loss val_perplexity
0 8.719033 6118.257553 8.449943 4674.806543
saving best model...
train loss: 7.032: 100% 40/40 [00:27<00:00, 2.63it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
with autocast():
valid loss: 6.549: 100% 79/79 [00:30<00:00, 3.87it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
with autocast():
train_loss train_perplexity val_loss val_perplexity
1 7.878206 2639.134687 6.417279 612.33488
saving best model...
unfreezing GPT2 entirely...
total_frozen_params=0
train loss: 5.926: 100% 40/40 [00:36<00:00, 2.90it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
with autocast():
valid loss: 5.549: 100% 79/79 [00:29<00:00, 3.89it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
with autocast():
train_loss train_perplexity val_loss val_perplexity
2 6.51366 674.289833 5.066973 158.693309
saving best model...
train loss: 5.099: 100% 40/40 [00:46<00:00, 2.95it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
with autocast():
valid loss: 4.972: 100% 79/79 [00:30<00:00, 3.78it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
with autocast():
train_loss train_perplexity val_loss val_perplexity
3 5.684741 294.341659 4.44958 85.591001
saving best model...
train loss: 5.077: 100% 40/40 [00:46<00:00, 2.91it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
with autocast():
valid loss: 4.502: 100% 79/79 [00:30<00:00, 3.71it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
with autocast():
train_loss train_perplexity val_loss val_perplexity
4 5.124145 168.030337 3.910129 49.905397
saving best model...
train loss: 4.912: 100% 40/40 [00:47<00:00, 2.95it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
with autocast():
valid loss: 4.136: 100% 79/79 [00:30<00:00, 3.77it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
with autocast():
train_loss train_perplexity val_loss val_perplexity
5 4.778204 118.890575 3.626201 37.569822
saving best model...
train loss: 4.081: 100% 40/40 [00:46<00:00, 2.86it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
with autocast():
valid loss: 3.990: 100% 79/79 [00:30<00:00, 3.75it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
with autocast():
train_loss train_perplexity val_loss val_perplexity
6 4.534204 93.149315 3.50339 33.227909
saving best model...
train loss: 4.425: 100% 40/40 [00:46<00:00, 2.87it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
with autocast():

valid loss: 3.931: 100% 79/79 [00:30<00:00, 3.97it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
with autocast():
train_loss train_perplexity val_loss val_perplexity
7 4.382705 80.054321 3.423486 30.676163
saving best model...

train loss: 4.496: 100% 40/40 [00:46<00:00, 2.95it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
with autocast():

valid loss: 3.865: 100% 79/79 [00:30<00:00, 3.90it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
with autocast():
train_loss train_perplexity val_loss val_perplexity
8 4.313307 74.687064 3.401876 30.020366
saving best model...

train loss: 4.098: 100% 40/40 [00:46<00:00, 2.97it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
with autocast():

valid loss: 3.862: 100% 79/79 [00:30<00:00, 3.66it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
with autocast():
train_loss train_perplexity val_loss val_perplexity
9 4.274336 71.832408 3.400049 29.965578
saving best model...

Metrics for the model trained on the distilled dataset:
train_loss train_perplexity val_loss val_perplexity
0 8.719033 6118.257553 8.449943 4674.806543
1 7.878206 2639.134687 6.417279 612.33488
2 6.51366 674.289833 5.066973 158.693309
3 5.684741 294.341659 4.44958 85.591001
4 5.124145 168.030337 3.910129 49.905397
5 4.778204 118.890575 3.626201 37.569822
6 4.534204 93.149315 3.50339 33.227909
7 4.382705 80.054321 3.423486 30.676163
8 4.313307 74.687064 3.401876 30.020366
9 4.274336 71.832408 3.400049 29.965578

--- COMPARISON ---
Full Dataset Training Metrics:
Empty DataFrame
Columns: [train_loss, train_perplexity, val_loss, val_perplexity]
Index: []

Distilled Dataset Training Metrics:
train_loss train_perplexity val_loss val_perplexity
0 8.719033 6118.257553 8.449943 4674.806543
1 7.878206 2639.134687 6.417279 612.33488
2 6.51366 674.289833 5.066973 158.693309
3 5.684741 294.341659 4.44958 85.591001
4 5.124145 168.030337 3.910129 49.905397
5 4.778204 118.890575 3.626201 37.569822
6 4.534204 93.149315 3.50339 33.227909
7 4.382705 80.054321 3.423486 30.676163
8 4.313307 74.687064 3.401876 30.020366
9 4.274336 71.832408 3.400049 29.965578