**Batch size 1, 300K rows of data, Bucket Size 262144, prefetch factor 400**

A screenshot of a computer

Description automatically generated

**Batch size 1, 10 rows of data, Bucket Size 262144, prefetch factor 400**

A graph of data on a computer

Description automatically generated with medium confidence A screenshot of a computer

Description automatically generated

[2023-08-25 02:59:31,948 WARNING] The batch will be filled until we reach 1, its size may exceed 1 tokens

[2023-08-25 03:00:03,450 WARNING] The batch will be filled until we reach 1, its size may exceed 1 tokens

[2023-08-25 03:00:37,147 WARNING] The batch will be filled until we reach 1, its size may exceed 1 tokens

/usr/lib/python3.10/multiprocessing/resource\_tracker.py:224: UserWarning: resource\_tracker: There appear to be 20 leaked semaphore objects to clean up at shutdown

warnings.warn('resource\_tracker: There appear to be %d '

^C

**Batch Size 1, 1 Row of data, Bucket Size 262144, prefetch factor 400**

A screenshot of a computer error

Description automatically generated

**Batch Size 1, 10 Row of data, Bucket Size 262144, prefetch factor 1**

A screen shot of a computer

Description automatically generated

A screen shot of a computer

Description automatically generated

[2023-08-25 03:25:36,614 INFO] Weighted corpora loaded so far:

\* cc-matrix-enzh: 131072

[2023-08-25 03:25:54,722 WARNING] The batch will be filled until we reach 1, its size may exceed 1 tokens

[2023-08-25 03:26:34,878 WARNING] The batch will be filled until we reach 1, its size may exceed 1 tokens

[2023-08-25 03:27:11,142 WARNING] The batch will be filled until we reach 1, its size may exceed 1 tokens

/usr/lib/python3.10/multiprocessing/resource\_tracker.py:224: UserWarning: resource\_tracker: There appear to be 20 leaked semaphore objects to clean up at shutdown

warnings.warn('resource\_tracker: There appear to be %d '

^C

**Batch Size 1, 10 Row of data, Bucket Size 8, prefetch factor 1**

A screenshot of a computer

Description automatically generated

[2023-08-25 03:30:09,193 INFO] Weighted corpora loaded so far:

\* cc-matrix-enzh: 2

Traceback (most recent call last):

File "/content/drive/MyDrive/OpenNMT-py/train.py", line 6, in <module>

main()

File "/content/drive/MyDrive/OpenNMT-py/onmt/bin/train.py", line 67, in main

train(opt)

File "/content/drive/MyDrive/OpenNMT-py/onmt/bin/train.py", line 52, in train

train\_process(opt, device\_id=0)

File "/content/drive/MyDrive/OpenNMT-py/onmt/train\_single.py", line 235, in main

trainer.train(

File "/content/drive/MyDrive/OpenNMT-py/onmt/trainer.py", line 301, in train

for i, (batches, normalization) in enumerate(self.\_accum\_batches(train\_iter)):

File "/content/drive/MyDrive/OpenNMT-py/onmt/trainer.py", line 231, in \_accum\_batches

for batch in iterator:

File "/content/drive/MyDrive/OpenNMT-py/onmt/inputters/inputter.py", line 33, in \_\_iter\_\_

for tensor\_batch in self.iterable:

File "/usr/local/lib/python3.10/dist-packages/torch/utils/data/dataloader.py", line 633, in \_\_next\_\_

data = self.\_next\_data()

File "/usr/local/lib/python3.10/dist-packages/torch/utils/data/dataloader.py", line 1345, in \_next\_data

return self.\_process\_data(data)

File "/usr/local/lib/python3.10/dist-packages/torch/utils/data/dataloader.py", line 1371, in \_process\_data

data.reraise()

File "/usr/local/lib/python3.10/dist-packages/torch/\_utils.py", line 644, in reraise

raise exception

UnboundLocalError: Caught UnboundLocalError in DataLoader worker process 3.

Original Traceback (most recent call last):

File "/content/drive/MyDrive/OpenNMT-py/onmt/inputters/dynamic\_iterator.py", line 85, in \_\_iter\_\_

item = next(iterator)

StopIteration

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

File "/content/drive/MyDrive/OpenNMT-py/onmt/inputters/dynamic\_iterator.py", line 89, in \_\_iter\_\_

item = next(iterator)

StopIteration

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

File "/usr/local/lib/python3.10/dist-packages/torch/utils/data/\_utils/worker.py", line 308, in \_worker\_loop

data = fetcher.fetch(index)

File "/usr/local/lib/python3.10/dist-packages/torch/utils/data/\_utils/fetch.py", line 41, in fetch

data = next(self.dataset\_iter)

File "/content/drive/MyDrive/OpenNMT-py/onmt/inputters/dynamic\_iterator.py", line 320, in \_\_iter\_\_

for bucket in self.\_bucketing():

File "/content/drive/MyDrive/OpenNMT-py/onmt/inputters/dynamic\_iterator.py", line 257, in \_bucketing

for ex in self.mixer:

File "/content/drive/MyDrive/OpenNMT-py/onmt/inputters/dynamic\_iterator.py", line 91, in \_\_iter\_\_

yield item

UnboundLocalError: local variable 'item' referenced before assignment

**Batch Size 1, 1 Row of data, Bucket Size 8, prefetch factor 1**

A screenshot of a computer

Description automatically generated

[2023-08-25 05:57:13,894 INFO] Weighted corpora loaded so far:

\* cc-matrix-enzh: 16

[2023-08-25 05:57:13,897 WARNING] The batch will be filled until we reach 1, its size may exceed 1 tokens

Traceback (most recent call last):

File "/content/drive/MyDrive/OpenNMT-py/train.py", line 6, in <module>

main()

File "/content/drive/MyDrive/OpenNMT-py/onmt/bin/train.py", line 67, in main

train(opt)

File "/content/drive/MyDrive/OpenNMT-py/onmt/bin/train.py", line 52, in train

train\_process(opt, device\_id=0)

File "/content/drive/MyDrive/OpenNMT-py/onmt/train\_single.py", line 235, in main

trainer.train(

File "/content/drive/MyDrive/OpenNMT-py/onmt/trainer.py", line 301, in train

for i, (batches, normalization) in enumerate(self.\_accum\_batches(train\_iter)):

File "/content/drive/MyDrive/OpenNMT-py/onmt/trainer.py", line 231, in \_accum\_batches

for batch in iterator:

File "/content/drive/MyDrive/OpenNMT-py/onmt/inputters/inputter.py", line 33, in \_\_iter\_\_

for tensor\_batch in self.iterable:

File "/usr/local/lib/python3.10/dist-packages/torch/utils/data/dataloader.py", line 633, in \_\_next\_\_

data = self.\_next\_data()

File "/usr/local/lib/python3.10/dist-packages/torch/utils/data/dataloader.py", line 1345, in \_next\_data

return self.\_process\_data(data)

File "/usr/local/lib/python3.10/dist-packages/torch/utils/data/dataloader.py", line 1371, in \_process\_data

data.reraise()

File "/usr/local/lib/python3.10/dist-packages/torch/\_utils.py", line 644, in reraise

raise exception

UnboundLocalError: Caught UnboundLocalError in DataLoader worker process 1.

Original Traceback (most recent call last):

File "/content/drive/MyDrive/OpenNMT-py/onmt/inputters/dynamic\_iterator.py", line 85, in \_\_iter\_\_

item = next(iterator)

StopIteration

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

File "/content/drive/MyDrive/OpenNMT-py/onmt/inputters/dynamic\_iterator.py", line 89, in \_\_iter\_\_

item = next(iterator)

StopIteration

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

File "/usr/local/lib/python3.10/dist-packages/torch/utils/data/\_utils/worker.py", line 308, in \_worker\_loop

data = fetcher.fetch(index)

File "/usr/local/lib/python3.10/dist-packages/torch/utils/data/\_utils/fetch.py", line 41, in fetch

data = next(self.dataset\_iter)

File "/content/drive/MyDrive/OpenNMT-py/onmt/inputters/dynamic\_iterator.py", line 320, in \_\_iter\_\_

for bucket in self.\_bucketing():

File "/content/drive/MyDrive/OpenNMT-py/onmt/inputters/dynamic\_iterator.py", line 257, in \_bucketing

for ex in self.mixer:

File "/content/drive/MyDrive/OpenNMT-py/onmt/inputters/dynamic\_iterator.py", line 91, in \_\_iter\_\_

yield item

UnboundLocalError: local variable 'item' referenced before assignment

**Batch Size 1, 10 Row of data, Bucket Size 1024, prefetch factor 2**

A screenshot of a computer

Description automatically generated

[2023-08-25 06:40:49,799 INFO] Weighted corpora loaded so far:

\* cc-matrix-enzh: 512

[2023-08-25 06:40:49,911 WARNING] The batch will be filled until we reach 1, its size may exceed 1 tokens

[2023-08-25 06:40:50,017 WARNING] The batch will be filled until we reach 1, its size may exceed 1 tokens

[2023-08-25 06:40:50,178 WARNING] The batch will be filled until we reach 1, its size may exceed 1 tokens

Traceback (most recent call last):

File "/content/drive/MyDrive/OpenNMT-py/train.py", line 6, in <module>

main()

File "/content/drive/MyDrive/OpenNMT-py/onmt/bin/train.py", line 67, in main

train(opt)

File "/content/drive/MyDrive/OpenNMT-py/onmt/bin/train.py", line 52, in train

train\_process(opt, device\_id=0)

File "/content/drive/MyDrive/OpenNMT-py/onmt/train\_single.py", line 235, in main

trainer.train(

File "/content/drive/MyDrive/OpenNMT-py/onmt/trainer.py", line 301, in train

for i, (batches, normalization) in enumerate(self.\_accum\_batches(train\_iter)):

File "/content/drive/MyDrive/OpenNMT-py/onmt/trainer.py", line 231, in \_accum\_batches

for batch in iterator:

File "/content/drive/MyDrive/OpenNMT-py/onmt/inputters/inputter.py", line 33, in \_\_iter\_\_

for tensor\_batch in self.iterable:

File "/usr/local/lib/python3.10/dist-packages/torch/utils/data/dataloader.py", line 633, in \_\_next\_\_

data = self.\_next\_data()

File "/usr/local/lib/python3.10/dist-packages/torch/utils/data/dataloader.py", line 1325, in \_next\_data

return self.\_process\_data(data)

File "/usr/local/lib/python3.10/dist-packages/torch/utils/data/dataloader.py", line 1371, in \_process\_data

data.reraise()

File "/usr/local/lib/python3.10/dist-packages/torch/\_utils.py", line 644, in reraise

raise exception

UnboundLocalError: Caught UnboundLocalError in DataLoader worker process 3.

Original Traceback (most recent call last):

File "/content/drive/MyDrive/OpenNMT-py/onmt/inputters/dynamic\_iterator.py", line 85, in \_\_iter\_\_

item = next(iterator)

StopIteration

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

File "/content/drive/MyDrive/OpenNMT-py/onmt/inputters/dynamic\_iterator.py", line 89, in \_\_iter\_\_

item = next(iterator)

StopIteration

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

File "/usr/local/lib/python3.10/dist-packages/torch/utils/data/\_utils/worker.py", line 308, in \_worker\_loop

data = fetcher.fetch(index)

File "/usr/local/lib/python3.10/dist-packages/torch/utils/data/\_utils/fetch.py", line 41, in fetch

data = next(self.dataset\_iter)

File "/content/drive/MyDrive/OpenNMT-py/onmt/inputters/dynamic\_iterator.py", line 320, in \_\_iter\_\_

for bucket in self.\_bucketing():

File "/content/drive/MyDrive/OpenNMT-py/onmt/inputters/dynamic\_iterator.py", line 257, in \_bucketing

for ex in self.mixer:

File "/content/drive/MyDrive/OpenNMT-py/onmt/inputters/dynamic\_iterator.py", line 91, in \_\_iter\_\_

yield item

UnboundLocalError: local variable 'item' referenced before assignment

**Batch Size 1, 1K Row of data, Bucket Size 1024, prefetch factor 2**

A screenshot of a computer

Description automatically generated