Homework 2 write-up

Enas Albasiri

In this assignment, I felt confident in applying the previous skills I learned to solve the given problem. I had no challenges or difficulties in Part 1 and 2 that requested preprocessing the data since now we have practiced similar tasks in the past. In part 3, we were asked to train the model, that was the most challenging part because reading the doc for Fairseq-training was very difficult and not very direct. I solved this by googling most of the arguments to see how people used them in codes.

```
Here is my WER: 17.00
My logging info:
Namespace(align_suffix=None, alignfile=None,
all_gather_list_size=16384, bf16=False, bpe=None,
checkpoint_shard_count=1, checkpoint_suffix='', cpu=False,
criterion='cross_entropy', dataset_impl='mmap', destdir='data-bin',
empty_cache_freq=0, fp16=False, fp16_init_scale=128,
fp16_no_flatten_grads=False, fp16_scale_tolerance=0.0,
fp16_scale_window=None, joined_dictionary=False, log_format=None,
log_interval=100, lr_scheduler='fixed', memory_efficient_bf16=False,
memory_efficient_fp16=False, min_loss_scale=0.0001,
model_parallel_size=1, no_progress_bar=False, nwordssrc=-1,
nwordstgt=-1, only_source=False, optimizer=None, padding_factor=8,
profile=False, quantization_config_path=None, scoring='bleu', seed=1,
source_lang='ice.g', srcdict=None, target_lang='ice.p',
task='translation', tensorboard_logdir=None, testpref='test',
tgtdict=None, threshold_loss_scale=None, thresholdsrc=2,
thresholdtgt=2, tokenizer='space', tpu=False, trainpref='train',
user_dir=None, validpref='dev', workers=1)
[ice.g] Dictionary: 40 types
[ice.g] train.ice.g: 800 sents, 5242 tokens, 0.0191% replaced by <unk>
[ice.g] Dictionary: 40 types
[ice.g] dev.ice.g: 100 sents, 634 tokens, 0.0% replaced by <unk>
[ice.g] Dictionary: 40 types
[ice.g] test.ice.g: 100 sents, 667 tokens, 0.0% replaced by <unk>
[ice.p] Dictionary: 64 types
[ice.p] train.ice.p: 800 sents, 5376 tokens, 0.0558% replaced by <unk>
[ice.p] Dictionary: 64 types
[ice.p] dev.ice.p: 100 sents, 652 tokens, 0.153% replaced by <unk>
```