EE738 Final Report
Cheryl M. Siy

After finishing the baseline code, the evaluation mode yielded a CER of 29.8%

In order to improve the performance, I have divided my implementation in 2 phases
1. Adding an LR scheduler and modifying hyperparameters
2. Fine tuning

Phase 1

In order to validate my hypothesis, I have modified some hyper parameters
- Epoch : 20
- Batch size : 32
- Drop-out : 0.3

The parameters are inspired by a work KoSpeech: Open-Source Toolkit for End-to-End Korean Speech Recognition [1]

Adding an LR scheduler

I have used the learning rate scheduler **ReduceLROnPlateau**. The main idea of this LR scheduler is that it reduces the learning rate when the metric has stopped improving. However, this produced a higher CER compared to the initial implementation.

 scheduler = optim.lr_scheduler.ReduceLROnPlateau(optimizer, mode ='min', patience = 10, factor = 0.1)

```
(myenv) C:\Users\Cheryl\Documents\SpeechRecog\speechrecog>python local.py --eval --initial_model LR.pt --
1.11.0
0.11.0
Model loaded. Number of parameters: 4884175
100%|                                                                    | 1500/1500
Character Error Rate is 26.20%
```

The documentation can be seen here:
https://pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.ReduceLROnPlateau.html

However, upon checking, the model without the Learning Rate Scheduler performed better. So, I later on fine tuned it to decrease CER.

```
(myenv) C:\Users\Cheryl\Documents\SpeechRecog\spee
1.11.0
0.11.0
Model loaded. Number of parameters: 4884175
100%|
Character Error Rate is 25.13%
```

Upon checking the train log files, the validation loss seems to be still decreasing, so I increase the epoch to 40 to further fine-tune. At around Epoch 23, I saw that the validation loss is not decreasing anymore.

For my final implementation, I have used training parameters of:
-   Epoch: 23
-   Batch-size: 32
-   Drop-out 0.3
To get a 24.77% CER.

```
(myenv) C:\Users\Cheryl\Documents\SpeechRecog\speechrecog>python local.py --eval --initial_mode
oLR23
1.11.0
0.11.0
Model loaded. Number of parameters: 4884175
100%|                                                              | 1500/1500
Character Error Rate is 24.77%
```

Evaluating the test set:
-   For evaluating the test set, I have changed the --val_list to "ks_test_rel.json"

References
[1] Kim, Soohwan, Seyoung Bae, and Cheolhwang Won. "Open-source toolkit for end-to-end Korean speech recognition." *Software Impacts* 7 (2021): 100054