Lab 7: Group work on projects

The goal of this lab is for you to make progress on your project, together as a group. You'll set goals and work towards them, and report what you got done, challenges you faced, and subsequent plans.

Group name: How Do you turn this on.

Group members present in lab today: Yi-Ting Yeh, Ting-Rui Chiang

1: Plan

1. What is your plan for today, and this week?
   Today: Synchronize the progress we have now, and decide what to do next.
   This week: 1) Evaluate the layerwise pruned model. 2) Apply head pruning on the layerwise prune model. 3) Analyze the result and decide the future direction.
2. How will each group member contribute towards this plan?
   Yi-Ting: Continue trying different settings of layer pruning.
   Ting-Rui: Evaluate the layerwise pruned model and apply head pruning on it.

2: Execution

1. What have you achieved today / this week? Was this more than you had planned to get done? If so, what do you think worked well?

   Today: We inspected the log of the fine-tuning process. We discussed whether the trajectories were reasonable. We also discussed what to do next, deciding the testing settings, which we elaborated in the next question.

   This week:

   (1) We continued to try different settings of training layerwise pruned ASR model with stochastic layer drop regularization and intermediate CTC loss. On fine-tuning pretrained ASR models, we found 1) for intermediate CTC, the performance is very sensitive to the choice of intermediate layer. Specifically, if we follow the original paper and apply CTC loss on layer 12 when fine-tuning 18 layer Transformer ASR models, the performance will drop significantly. However, if we choose layer 16 instead, then we can keep the original performance to a certain degree. We think it is because the representation learned with CTC will be very different to the representation from the learned intermediate layer.

(2) We evaluated the performance of a layerwise pruned model.
    i. The model was fine-tuned with stochastic layer drop regularization and an intermediate CTC loss at layer 16: We evaluated the model with two settings: 1) Truncating layers after 16, i.e. removing layer 16, 17, 18. The WER was 0.30 2) Dropping about 1 / 3 of the layers uniformly, i.e. dropping layer 2, 5, 8, 11, 14, 17. The WER was 0.29.
    ii. The model only trained with stochastic layer drop regularization: 1) Dropping out ⅓ layers uniformly gives us WER 0.15 2) Dropping out ½ layers uniformly gives us WER 0.22. 3) Dropping out ⅔ layers gives us WER 0.50
(3) We applied head pruning over the above models in the above two settings. We have finished the evaluation script. We are still waiting for some results.

2. Was there anything you had hoped to achieve, but did not? What happened? How did you work to resolve these challenges?

   We hoped we would finish head-pruning over the layerwise pruned models this week. However, fine-tuning with stochastic layer drop and intermediate CTC was much more unstable than we had expected. The performance was also not as expected either. We may need to try more hyper-parameter configurations. Since adding intermediate CTC loss makes the fine-tuning process unstable the most, we may use only stochastic layer drop.

3. What were the contributions of each group member towards all of the above?
   Yi-Ting: Layer pruning.
   Ting-Rui: Evaluating the layerwise pruned models, and applying head pruning on them.

3: Next steps

1. Are you making sufficient progress towards completing your final project? Explain why or why not. If not, please report how you plan to change the scope and/or focus of your project accordingly.

   In this week, we found intermediate CTC loss might not be applicable in the finetuning setting. Due to the time and computing resource constraints, we might not want to train the ASR model from scratch. On the other hand, only using stochastic layer drop gives us surprisingly good results. We can remove half of the layers and still obtain reasonable results. While we will still try to figure out a way to do intermediate CTC on the fine-tuning regime, we plan to explore more in the direction of stochastic layer drop.

2. Based on your work today / this week, and your answer to (1), what are your group's planned next steps?
    (1) We will continue tuning the hyper-parameters for layerwise pruning.
    (2) We may need to come up with a more sophisticated way to decide which layers to drop. Since we may not be able to reproduce the results in the original paper by fine-tuning, their layer-dropping strategy may not be applicable in our case.
3. How will each group member contribute towards those steps?

    Yi-Ting: Exploring other strategies of fineuning layewise pruned models
    Ting-Rui: Layer-dropping strategies.