Lab 8: 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. Analyze current experiment results. and decide what to do next.
   This week: We would like to finish the exploration of intermediate CTC. If it doesn't work well this week we will give up and use only dynamic depth to prune our model.
2. How will each group member contribute towards this plan?
   Yi-Ting: Fix the bug in our intermediate CTC implementation in ESPNet2. Train new models with the correct version of the code.
   Ting-Rui: Evaluate the models in different settings.

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 collected and discussed our current experiment results. We decided to do the last trial with intermediate CTC. If we still can't make it work, we would use only dynamic depth to fine-tune the layer-prune model.

   This week:

   (1) We re-investigated our implementation of intermediate CTC loss in ESPNet2. We found when porting the feature from ESPNet1, we forgot to add an additional normalization layer before computing the intermediate CTC loss. We fixed this bug and now could train the model with intermediate CTC successfully.
   (2) For the model fine-tuned with dynamic depth and intermediate CTC at 6th and 12th layers, we test its performance when using only the first 6 layers and 12 layers. The model with the first 6 layers has WER 0.34 and WER 0.25 with the

first 12 layers. The full model also achieves WER 0.25. We also observe the overfitting phenomenon when using intermediate CTC. The validation accuracy keeps decreasing during the fine-tuning process. If we simply choose the checkpoints with the best validation accuracy and use its first 6 or 12 layers, we will obtain WER 0.74 and 0.42 respectively. Compared to the model trained with only dynamic depth, the full model achieved WER 0.06 but had WER 0.22 and 0.50 when we uniformly remove 9 layers and 12 layers from the model. The intermediate CTC is actually effective in aggressively pruning the model. Therefore, the way to deal with overfitting when using intermediate CTC loss will be the next topic.

(3) We found a bug in the evaluation script. In the previous version, the script evaluated the models in the transformer mode. We fixed the problem, so now the script will evaluate the performance of the models in the CTC mode.

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

   We still could not find the best training setting for intermediate CTC. As described above, the performance was not as what we had expected. We will continue trying different training strategies. If we still cannot find a solution, we may not use intermediate CTC in the end.

3. What were the contributions of each group member towards all of the above?
   Yi-Ting: Identify the bug in the implementation of intermediate CTC. Train and test new models with the fixed version of the intermediate CTC layer.
   Ting-Rui: Identify the bug in the evaluation script, and evaluate the scripts.

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.

Yes, we have identified the key issue in our implementation of intermediate CTC loss and finally are able to use it to prune our model.

2. Based on your work today / this week, and your answer to (1), what are your group's planned next steps?

   (1) We plan to continue trying different settings for using intermediate CTC loss. We will focus on training with stochastic layer dropping if intermediate CTC loss still doesn't work.

(2) We will implement an algorithm that selects the layers to drop.
(3) We will also re-evaluate the previous model with layer pruning with our fixed evaluated script.

3. How will each group member contribute towards those steps?

Yi-Ting: Try more sophisticated fine-tuning strategies / layer-dropping stretegeies of layer pruning models. Working on solving the overfitting issue.
Ting-Rui: Implement the algorithm that selects the layers to drop and re-evaluate the previous models with the fixed script.