Tutoring

Tuesday, November 2, 2021 11:27 AM

- Use Speechbrain
- Introduction to SpeechBrain tutorial: https://colab.research.google.com/drive/12bg3aUdr9mTfOGqcB5pSMABoIKPgiwcM?usp=sharing
- Use TIMIT recipe
- Run each configuration at least 2 times.
- Check https://speechbrain.readthedocs.io/en/latest/experiment.html for an example on how to run a recipe.
- Use the TIMIT/ASR/seq2seq/hparams/train.yaml hyperparameter file (no wav2vec 2.0). The corresponding training script is TIMIT/ASR/seq2seq/train.py
- `TIMIT under Aalto machines: `/m/teamwork/t40511_asr/c/timit/`
  - Alternatively find TIMIT here: https://github.com/philipperemy/timit
- Steps
  1. First run the default TIMIT recipe in order to get acquainted with how the toolkit works. You can also follow the tutorial linked above.
  2. Next run the default TIMIT recipe and change the `sorting` attribute to `descending`.
  3. Run the above experiments as many times required in order to get a stable average PER on the test set.
  4. To get the best results, try out some different hyperparameters (learning rate, ctc weight, dnn size, beam size and maybe others).
  5. Next, subclass `speechbrain.dataio.dataset.DynamicItemDataset` and change the `filtered_sorted` method so that it will also be able to sort based on new attributes.
     - To achieve that you can extend the number of input arguments.
     - An extra argument which you will probably need is a hash map (e.g. a dictionary) mapping from every training example to a single score.
     - Then you can use this hash map in order to create the `filtered_sorted_ids` variable (a list) which should then be returned by the function.
  6. In the above step, you will practically be creating a scoring function which calculates the easiness of each example.
  7. Scoring functions. Choose one of the following:
     i. Two loss-based scoring functions.
     ii. One metric-based scoring function (harder).
     iii. One scoring function (whatever you want) and the amount of training data increase on every epoch. E.g. for the first 5 epochs, train with 20% of the easiest examples, for the next 5 epochs train with the 35%, then 50%, e.t.c. This may lead to overfitting so you should be very careful if you choose this method. It is, however, the method that is closest to the way humans learn.
  8. Loss based scoring function:
     i. Seq2Seq loss -> Calculate the seq2seq loss for each training EXAMPLE (not batch).
     ii. Seq2Seq+CTC loss -> Combination of seq2seq and ctc loss.
  9. Metric based scoring function:
     i. WER based -> Calculate WER of each training example and sort based on that value.
     ii. PER based -> Same but with PER (probably easier).
     iii. WER/PER based with confidences -> Same as above but also normalize the values by using the predictions' confidences.
- Division of work:
  - Everyone should get familiar with speechbrain, so the first 2 steps should be performed by the whole team.
  - The hyperparameter tuning phase, could be split among the team members.
  - After hyperparameter-tuning is complete, use the best hyperparameters in order to train your models with Curriculum Learning.