

No Maintenance Intended 

 TensorFlow

1.15.3

 TensorFlow

2.3

DeepSpeech2 Model

Overview

This is an implementation of the [DeepSpeech2](#) model. Current implementation is based on the code from the authors' [DeepSpeech code](#) and the implementation in the [MLPerf Repo](#).

DeepSpeech2 is an end-to-end deep neural network for automatic speech recognition (ASR). It consists of 2 convolutional layers, 5 bidirectional RNN layers and a fully connected layer. The feature in use is linear spectrogram extracted from audio input. The network uses Connectionist Temporal Classification [CTC](#) as the loss function.

Dataset

The [OpenSLR LibriSpeech Corpus](#) are used for model training and evaluation.

The training data is a combination of train-clean-100 and train-clean-360 (~130k examples in total). The validation set is dev-clean which has 2.7K lines. The download script will preprocess the data into three columns: wav_filename, wav_filesize, transcript. data/dataset.py will parse the csv file and build a tf.data.Dataset object to feed data. Within each epoch (except for the first if sortagrad is enabled), the training data will be shuffled batch-wise.

Running Code

Configure Python path

Add the top-level /models folder to the Python path with the command:

```
export PYTHONPATH="$PYTHONPATH:/path/to/models"
```

Install dependencies

First install shared dependencies before running the code. Issue the following command:

```
pip3 install -r requirements.txt
```

or

```
pip install -r requirements.txt
```

Run each step individually

Download and preprocess dataset

To download the dataset, issue the following command:

```
python data/download.py
```

Arguments:

- `--data_dir` : Directory where to download and save the preprocessed data. By default, it is `/tmp/librispeech_data`.

Use the `--help` or `-h` flag to get a full list of possible arguments.

Train and evaluate model

To train and evaluate the model, issue the following command:

```
python deep_speech.py
```

Arguments:

- `--model_dir` : Directory to save model training checkpoints. By default, it is `/tmp/deep_speech_model/`.
- `--train_data_dir` : Directory of the training dataset.
- `--eval_data_dir` : Directory of the evaluation dataset.
- `--num_gpus` : Number of GPUs to use (specify -1 if you want to use all available GPUs).

There are other arguments about DeepSpeech2 model and training/evaluation process. Use the `--help` or `-h` flag to get a full list of possible arguments with detailed descriptions.

Run the benchmark

A shell script [run_deep_speech.sh](#) is provided to run the whole pipeline with default parameters. Issue the following command to run the benchmark:

```
sh run_deep_speech.sh
```

Note by default, the training dataset in the benchmark include train-clean-100, train-clean-360 and train-other-500, and the evaluation dataset include dev-clean and dev-other.