

NYCU-IAII-ML2024 - Recurrent Neural Networks

蒲品憶 313581038 智能系統 碩一

前處理

- ▶ 先將音訊依照題目要求轉換成 **16 kHz sampling, signed-integer, 16 bits**
 - ▶ Linux系統，過程中使用sox library
 - ▶ 如果不想使用linux系統，也可以直接從網路上下載對應的Tool，進行轉換

使用sox library 方法 ----->

```
transform_wav_to_16K() {
    local SOURCE_DIR="$1"
    local DEST_DIR="$2"
    local fill_length="$3"

    rm -rf "$DEST_DIR"
    mkdir -p "$DEST_DIR"

    for file in "$SOURCE_DIR"/*.wav
    do
        filename=$(basename "$file")

        # parse index from filename
        # original filename regex: number.wav
        index=$(echo "$filename" | grep -o -E '[0-9]+' | head -1)

        index_number=$(printf "%0${fill_length}d" $index)
        new_filename="${index_number}.wav"
        sox "$file" -r 16000 -e signed-integer -b 16 "$DEST_DIR/$new_filename"
    done
}

transform_wav_to_16K "datasets/test" "datasets/transform/test" 0
transform_wav_to_16K "datasets/train" "datasets/transform/train" 0
```

前處理

- ▶ 將音訊、音訊檔案路徑和正確答案做結合
- ▶ 將資料上傳至hugging face 方便管理
- ▶ 幫助在後續讀檔、訓練時容易查找和使用

```
file_name,transcription
data/train/1.wav,li be e mih kiann lan lan san san long be tsiau tsng
data/train/10.wav,li tsit ma tai tsi na m bian liong tsa ka gua kong
data/train/1000.wav,a nia io a tia sioh io lai io khi a ti a u khun bo
data/train/1001.wav,gua i king bu tshing tsho tsit kiann tai tsi ah to sia li lai guan tau ka gua thong ti
data/train/1002.wav,ti thioh san tshi giam soo hok bu e kueh bing tik gian huat tshut sann tuann sik e ti tiau sai jio tsh
data/train/1003.wav,gua na e tsiah sue tak pian long tu tloh nua tho hue gua na e tsiah phaunn un tak pian sing li long si
data/train/1004.wav,tshin ke kin a jit e lai tsia
data/train/1006.wav,tsha put to si pai lak am kau le pai soo i khi phenn oo kim mng e ping iu ah si beh tshut kok e ping i
data/train/1007.wav,lan m thang ui huan tui lai huan tui tsi iau kuai kuai a thiann tsing hu e ue tsai i bu tsai i pinn
data/train/1008.wav,lin fans kin a jit na bo lai
data/train/1009.wav,tsa boo lang khia su khu ta khah ho khia tua tsiah e e sak be tin tang
data/train/1011.wav,peh pah khang ji phinn gua kau tsing tshik kah ban bah tah tsua
data/train/1012.wav,beh tai pak tshong siann mih
data/train/1013.wav,ke khi ap jip lai lam hai liau au e tsing kiong tsha put to si kin a jit am si ah si bin a tsai e piar
data/train/1014.wav,kha khai si leh thiann ah m tsai iann penn inn kiann e kau kiann be kau
data/train/1015.wav,tsit ma thoo ma tooh lu lai lu tinn m ho tsiah siunn tse
data/train/1016.wav,ho ah lah mai tsha lan koh lai khi khuann iau u khah sui e bo a bu an ne ka kong
data/train/1017.wav,li kiann siann hue lah
data/train/1018.wav,tsioo koo le too honnh ai suan tshit tsap goo pha sian tooh i siong e tui sin the khah ho
data/train/1019.wav,long tsong hing hing tng khi tso sin the kuai sun e loo le
data/train/102.wav,hann
data/train/1020.wav,ti hng hng u thiann tloh kui e hu jin lang tshinnh tshinnh e siann
data/train/1022.wav,tsi goo seh lin long be
data/train/1023.wav,hiah ni hiau
data/train/1025.wav,li beh tsham mi sooh tsin tsing ai sing tshi tsit e a kiam tsiann u e kiam tsiann bo siann kang khuann
data/train/1026.wav,in kong me ni kui ni long be tng khi hua lian
data/train/1028.wav,Ehi pang king ti leh hit ping lah
data/train/1029.wav,a pah e thau mng pun te oo sim sim tsiah bo kui ni suah long peh ah
data/train/103.wav,an ne e hue tshau m tsai siann to gua tse ai tshing sui e tsau gin a leh m koh kua beh kui gueh jit lor
data/train/1030.wav,tui oh im gak e lang lai kong e hiau puann phi a nooh si kai ki pun e tai tsi
data/train/1031.wav,tsiah e loh e he sinn mia thun tsiah be loh e thau thau a ka inn a tshang leh te a phah sng beh tsah t
data/train/1034.wav,tse iah he
data/train/1035.wav,a ing khuann i hiah sainn nai khui to tau koo tsiann hoo in tshit tho khah huann hi leh
data/train/1036.wav,tsue kin kue liau an tsuann tsit ma sin the u khah ho bo
data/train/1037.wav,tha looh sang tshian li tshui se giah hoo kuan
data/train/1038.wav,to sia honnh
data/train/1039.wav,lian kua e hong huat pah pah khuann u e si ka ti tsit e tsau khi kha la oo khe lian kua
```

Datasets: Amandapu/RNN_Taiwan

Modalities: Audio Text Formats: parquet Size: 1K - 10K Libraries: Datasets pandas Croissant +1

Dataset card Viewer Files and versions Community Settings

Dataset Viewer

Split (3)
train · 2.65k rows

Search this dataset

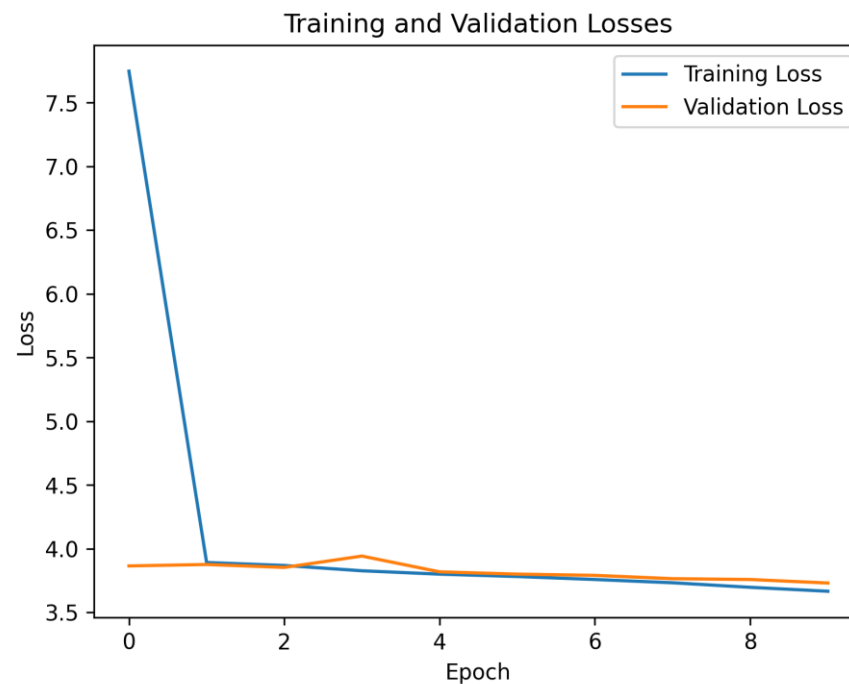
audio · duration (s) transcription · string · lengths

0.2 69.8	1 753
▶ 0:00 / 0:03	li be e mih kiann lan lan san san long be tsiau tsng
▶ 0:00 / 0:03	li tsit ma tai tsi na m bian liong tsa ka gua kong
▶ 0:00 / 0:05	a nia io a tia sioh io lai io khi a ti a u khun bo
▶ 0:00 / 0:05	gua i king bu tshing tsho tsit kiann tai tsi ah to sia li lai guan tau ka gua thong ti
▶ 0:00 / 0:08	ti thioh san tshi giam soo hok bu e kueh bing tik gian huat tshut sann tuann sik e ti tiau sai jio tshu li he thong
▶ 0:00 / 0:08	gua na e tsiah sue tak pian long tu tloh nua tho hue gua na e tsiah

◀ Previous 1 2 3 ... 27 Next ▶

訓練模型 1

- ▶ 原先使用老師統提供的DeepSpeech 2參考腳本進行修改
 - ▶ 效果大概是這樣，loss有下降
 - ▶ 問題是WER跟CER計算上有些誤會沒算出來



訓練模型 2

- ▶ 後來使用Espnet model
 - ▶ 雖然這次WER跟CER計算上有算出來但效果沒有好..... 所以轉戰!

訓練模型 3

► Whisper model - **Whisper-Finetuning**

- Medium
- Large

Size	Parameters	English-only model	Multilingual model	Required VRAM	Relative speed
tiny	39 M	<code>tiny.en</code>	<code>tiny</code>	~1 GB	~10x
base	74 M	<code>base.en</code>	<code>base</code>	~1 GB	~7x
small	244 M	<code>small.en</code>	<code>small</code>	~2 GB	~4x
medium	769 M	<code>medium.en</code>	<code>medium</code>	~5 GB	~2x
large	1550 M	N/A	<code>large</code>	~10 GB	1x
turbo	809 M	N/A	<code>turbo</code>	~6 GB	~8x

訓練模型 3

- ▶ Whisper model - **Whisper-Finetuning**
 - ▶ Large - 設定 “5000” **MAX_STEPS**
 - ▶ 最後跑到100 steps 硬體條件不足 memory 空間不夠

whisper-medium-tawiwanease-asr-200.csv

Complete · 7d ago

2.48971


2.94174



訓練模型 3

- ▶ Whisper model - **Whisper-Finetuning**
 - ▶ Medium - 設定 “5000” **MAX_STEPS**
 - ▶ 最後跑到1000 steps 硬體條件不足 memory 空間不夠
 - ▶ 注意：訓練模型設定 checkpoint很重要!!!
 - ▶ 如果有跑完，我覺得可以誤差值小於 1

名稱	修改日期	類型
 whisper-medium-taiwanese-asr-200	2024/12/19 下午 05:31	Microsoft Ex
 whisper-medium-taiwanese-asr-400	2024/12/19 下午 05:25	Microsoft Ex
 whisper-medium-taiwanese-asr-600	2024/12/19 下午 05:11	Microsoft Ex
 whisper-medium-taiwanese-asr-800	2024/12/19 下午 04:44	Microsoft Ex

	whisper-medium-taiwanese-asr.csv Complete · 7d ago	1.63374	1.64077
---	--	----------------	----------------