

Embedded system of Transformer-based TTS

Sian-Yi Chen

Advisors : Tay-Jyi Lin and Chingwei Yeh

Outline

Action item

- 完成 Transformer-based TTS 的嵌入式系統 (尚未完成)

Status report

- 先前進度
 - 完成使用C library json-c讀取input的json檔功能，並解決因無法直接在windows安裝、gcc編譯位元以及json-c語法問題
- TTS實作進度 (進行中)
 - 整理並記錄python版本所有執行的程式碼以及傳遞的參數，以下為大致的程式流程
 1. 讀命令列參數、模型超參數
 2. 讀input文字檔(json)
 3. 導入Transformer模型
 4. 根據input檔的id載入特徵參數
 5. 計算
 - 目前在處理的問題
 - 上述流程的第三步驟：原先原始碼編輯器使用中斷點無法進到Transformer底層，目前已更改完設定，並記錄底層執行過程所有會使用到的程式碼
 - 上述流程的第四步驟：根據input的id載入的特徵參數檔(.ark)為kaldi特有的binary儲存格式，使用一般的16進位編輯器無法辨別內容，學長建議可以看python中將ark檔轉換成numpy array的函數kaldiio (IO的部分)，然後在github中kaldi是使用c++寫的，或許部份程式碼可以抓來使用

附錄

```
input_test.py > ...
1 import kaldio
2
3 filepath = "exp/wei_hand_train_pytorch_train_pytorch_transformer+spkemb.tts1_en_zh_TMM1/spk_xvector.ark:9"
4 numpy_array = kaldio.load_mat(filepath)
5 print("numpy_array = ", numpy_array)
```

```
(tts) D:\Work_Space\off_line_speech_converter\paper_speech_resynthesis\WCC_2_C\WCC\tts>python input_test.py
numpy_array = [-8.67599845e-01 -2.28993368e+00 -2.30723119e+00 -2.50504756e+00
-1.40985012e+00 -2.13306499e+00 1.36214602e+00 -1.39892315e-02
2.85689449e+00 -5.76468647e-01 -2.04895186e+00 -3.05196762e+00
-3.81800294e+00 -3.08807278e+00 4.86751139e-01 3.05085540e+00
2.15426445e+00 -7.54291654e-01 1.67432845e+00 -2.36958623e+00
2.71533012e+00 6.48296535e-01 3.27817035e+00 -2.07637936e-01
-3.81869674e-01 1.76336527e+00 2.68878627e+00 1.19575822e+00
-1.71920612e-01 5.71759939e-01 6.37777328e-01 1.06703734e+00
-7.44350910e-01 2.79604077e-01 -4.30759335e+00 9.56195652e-01
1.93125629e+00 3.66337150e-02 -1.14268589e+00 2.13347030e+00
4.23614359e+00 -3.76055539e-01 -8.37063342e-02 1.37201393e+00
-1.67328787e+00 1.39213729e+00 -1.45180643e-01 2.40875626e+00
-9.78974044e-01 -6.66944742e-01 2.92323470e+00 -3.00652766e+00
-3.35206461e+00 -1.48611426e-01 1.94404259e-01 -2.44603491e+00
-8.94280195e-01 -1.37164950e+00 2.48773456e-01 2.74760389e+00
1.07200432e+00 -2.71635222e+00 2.44388747e+00 1.26828957e+00]
```

↓ 16位元編輯器

其中的資料總共有 128*4 共 512 筆↑

```
exp > wei_hand_train_pytorch_train_pytorch_transformer+spkemb.tts1_en_zh_TMM1 > spk_xvector.ark
```

00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F	Decoded Text
77 65 69 5F 68 61 6E 64 20 00 42 46 56 20 04 00	w e i _ h a n d . B F V . .
02 00 00 06 1B 5E BF 46 8E 12 C0 AD A9 13 C0 B3 ^ . F
52 20 C0 F8 75 B4 BF 23 84 08 C0 CD 5A AE 3F 17	R . . u . . # Z . ? .
33 65 BC 5C D7 36 40 73 93 13 BF 07 22 03 C0 70	3 e \ . 6 @ s " . . p
53 43 C0 29 5A 74 C0 FC A2 45 C0 72 37 F9 3E 37	S C .) Z t . . . E . r 7 . > 7
41 43 40 78 DF 09 40 42 19 41 BF 65 50 D6 3F 4D	A C @ x . . . @ B . A . e P . ? M
A7 17 C0 F8 C7 2D 40 C3 F6 25 3F 8B CD 51 40 0A - @ . . % ? . . Q @
9F 54 BE 6C 84 C3 BE F4 B5 E1 3F 13 15 2C 40 9B	. T . l ? . . , @ .
0E 99 3F F5 0B 30 BE DC 5E 12 3F 60 45 23 3F AE	. . ? . . 0 . . ^ . ? ` E # ? .
94 88 3F C8 8D 3E BF 44 28 8F 3E CE D7 89 C0 3D	. . ? . . > . D (. > . . . =
C9 74 3F 68 33 F7 3F 3C 0D 16 3D 88 43 92 BF C7	. t ? h 3 . ? < . . = . C . .
8A 08 40 7D 8E 87 40 5A 8A C0 BE 3A 6E AB BD 27	. . @ } . . @ Z . . . : n . . '
9E AF 3F 4C 2E D6 BF 8E 31 B2 3F 3C AA 14 BE 10	. . ? L 1 . ? < . . .
29 1A 40 0B 9E 7A BF E4 BC 2A BF 47 16 3B 40 F3) . @ . . z . . . * . G . ; @ .
6A 40 C0 3A 88 56 C0 98 2D 18 BE E9 11 47 3E D6	j @ . . : . V . . - . . . G > .
8B 1C C0 8C EF 64 BF 36 92 AF BF 78 BE 7E 3E BE d . 6 . . . x . ~ > .
D8 2F 40 70 37 89 3F B7 D8 2D C0 A7 68 1C 40 50	. / @ p 7 . ? . . - . . h . @ P
57 A2 3F 8F 21 BE BF C8 01 F7 BF D2 66 51 3E 7C	W . ? . ! f Q >
F9 02 C0 31 F6 44 3F 2B 0E A6 BE 47 25 FE BF 59	. . . 1 . D ? + . . . G % . . Y