

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

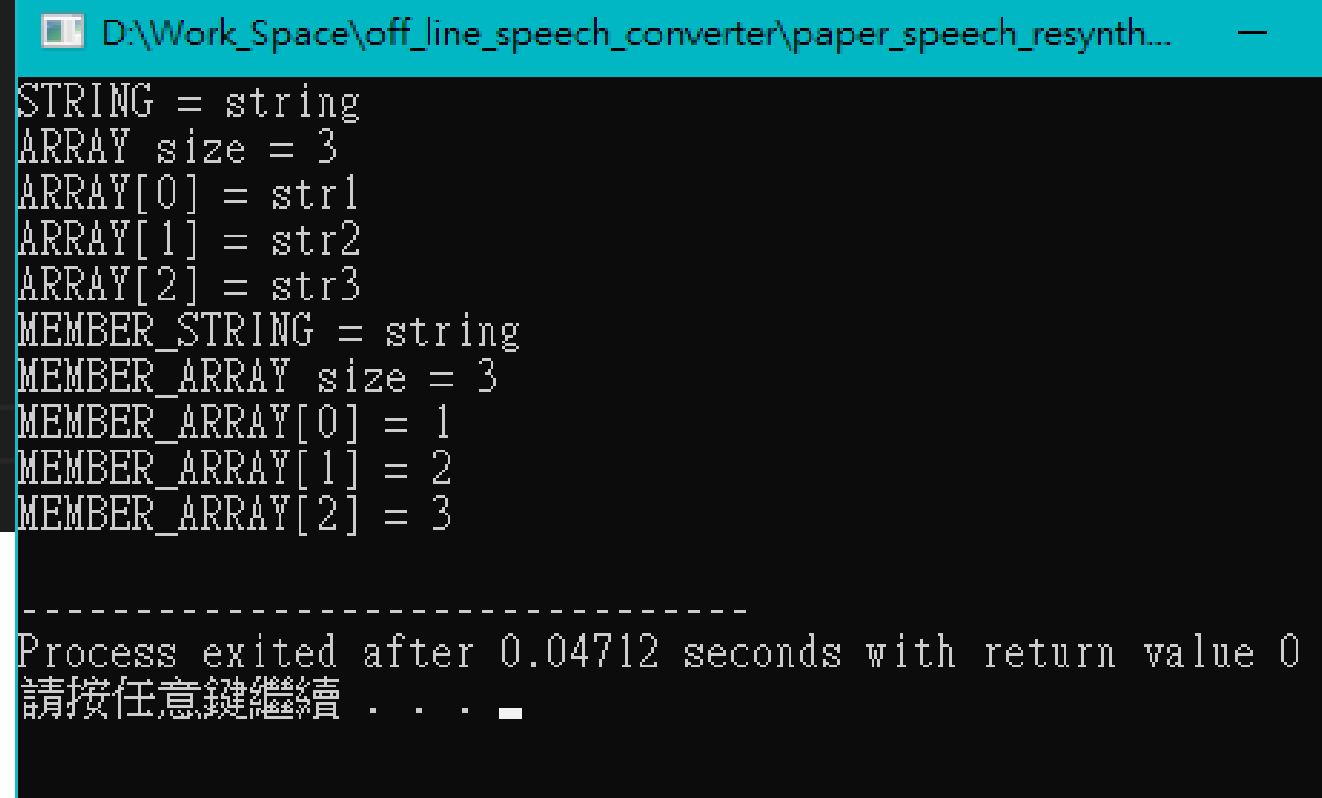
- 先前進度
 - 將TTS獨立從Linux取出至Windows能執行，並逐行確認程式功能。
 - 原先想仿照python中的argparse函數達到自由引用命令列變數，但最後決定直接設定為固定變數
- TTS實作進度 (進行中)
 - 使用json-c讀取input的json檔功能(完成)
 - 實現TTS讀取input的json檔功能的過程中，因ASR與TTS使用json作為數據交換的格式，因此可能無法省略，在c語言中沒有支援json格式，需要安裝額外的library(選擇了json-c)，但在實行中遇到不少問題
 1. 因json-c無法直接在windows系統上安裝，最後找到vcpkg套件類似pip或apt-get的套件管理器，支援在windows下安裝json-c
 2. 在安裝時遇到GCC編譯結果位元與目標不符，因此windows無法讀取，後續安裝能編x86-64的GCC解決
 3. 完成一版可以讀取簡單json格式的程式，但目標json格式更為複雜，在讀取多層結構出錯，最後找到教程解決找不到正確函數的窘境
 - 定義Transformer模型(進行中)
 - 理解espnet中許多自定義函數、刪除沒用到的程式，並執行確保無誤
 - 目前正在寫 embedding 的部分

附錄

這是目前完成 JSON檔結構 與 印出結果圖



```
1  {
2      "STRING": "string",
3      "ARRAY": ["str1", "str2", "str3"],
4      "OBJECT": {
5          "MEMBER_STRING": "string",
6          "MEMBER_ARRAY": [1, 2, 3]
7      }
8  }
```



```
D:\Work_Space\off_line_speech_converter\paper_speech_resynth...
STRING = string
ARRAY size = 3
ARRAY[0] = str1
ARRAY[1] = str2
ARRAY[2] = str3
MEMBER_STRING = string
MEMBER_ARRAY size = 3
MEMBER_ARRAY[0] = 1
MEMBER_ARRAY[1] = 2
MEMBER_ARRAY[2] = 3

-----
Process exited after 0.04712 seconds with return value 0
請按任意鍵繼續 . . .
```

但是我的目標格式長這樣，總共有 20 筆資料，目前取其中一筆(S0301)

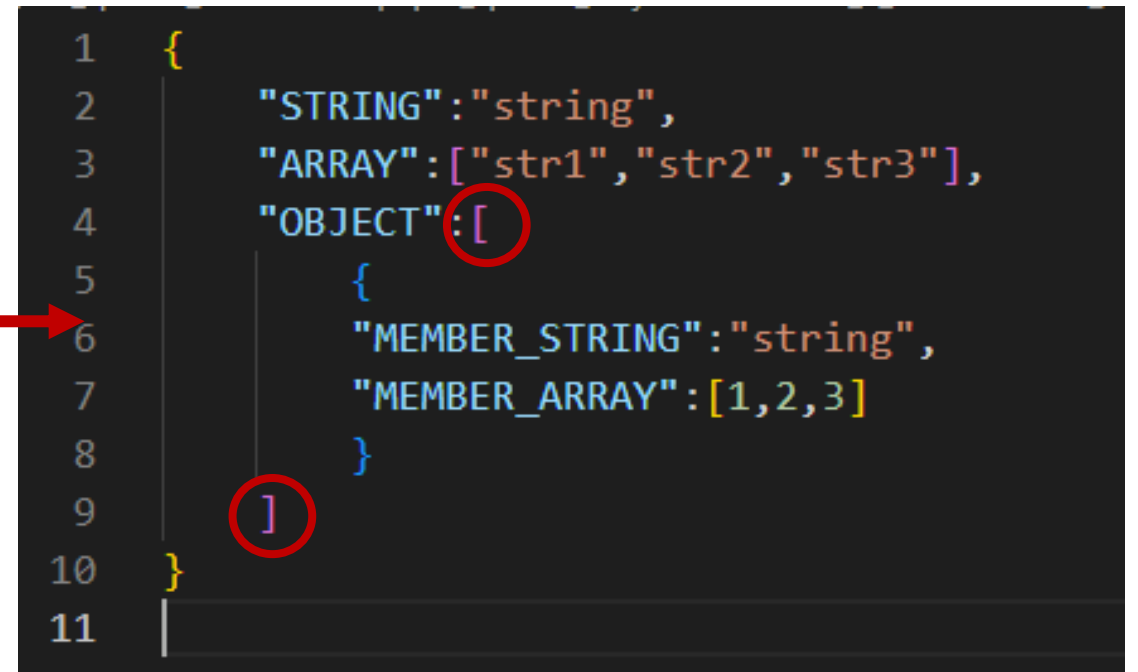
```
test2.json  test.json  x  example1.json
D: > Work_Space > off_line_speech_converter > paper_speech_resynthesis > VCC_2_C > VCC > tts_C > json > test.json > ...

1  {
2    "utts": {
3      "ming_24_wei_hand_S0301": {
4        "input": [
5          {
6            "feat": "exp/wei_hand_train_pytorch_train_pytorch_transformer+spkemb.tts1_en_zh_TMM1/spk_xvector.ark:9",
7            "name": "input1",
8            "shape": [512]
9          }
10         ],
11        "output": [
12          {
13            "name": "target1",
14            "shape": [18,337],
15            "text": "x iao3 p eng2 iou3 m en5 z ai4 k an4 m a5 i4 b an1 j ia1",
16            "token": "x iao3 p eng2 iou3 m en5 z ai4 k an4 m a5 i4 b an1 j ia1",
17            "tokenid": "333 174 248 133 216 228 131 334 86 226 93 228 82 151 113 90 225 153"
18          }
19         ],
20        "utt2spk": "wei_hand"
21      }
22    }
23  }
24
```

因此將第一頁的結構加了以下變化(紅框)



```
1 {
2   "STRING": "string",
3   "ARRAY": ["str1", "str2", "str3"],
4   "OBJECT": {
5     "MEMBER_STRING": "string",
6     "MEMBER_ARRAY": [1, 2, 3]
7   }
8 }
```



```
1 {
2   "STRING": "string",
3   "ARRAY": ["str1", "str2", "str3"],
4   "OBJECT": [
5     {
6       "MEMBER_STRING": "string",
7       "MEMBER_ARRAY": [1, 2, 3]
8     }
9   ]
10 }
11
```

下面圖為目標結構



```
{
  "utts": {
    "ming_24_we_i hand_S0301": {
      "input": [
        {
          "feat": "exp/wei_hand_t",
          "name": "input1",
          "shape": [512]
        }
      ],
    }
  }
}
```

但如果新增了 “[]” 一層結構，會導致程式抓不到該物件

```
1  {
2      "STRING": "string",
3      "ARRAY": ["str1", "str2", "str3"],
4      "OBJECT": [
5          {
6              "MEMBER_STRING": "string",
7              "MEMBER_ARRAY": [1, 2, 3]
8          }
9      ]
10 }
11
```

```
D:\Work_Space\off_line_spe
STRING = string
ARRAY size = 3
ARRAY[0] = str1
ARRAY[1] = str2
ARRAY[2] = str3
MEMBER_STRING = string
MEMBER_ARRAY size = 3
MEMBER_ARRAY[0] = 1
MEMBER_ARRAY[1] = 2
MEMBER_ARRAY[2] = 3
```

```
D:\Work_Space\off_line_speech_converter\
STRING = string
ARRAY size = 3
ARRAY[0] = str1
ARRAY[1] = str2
ARRAY[2] = str3
Cannot get MEMBER_STRING object
Fail to read a json file
```