Disentangled representation online resource

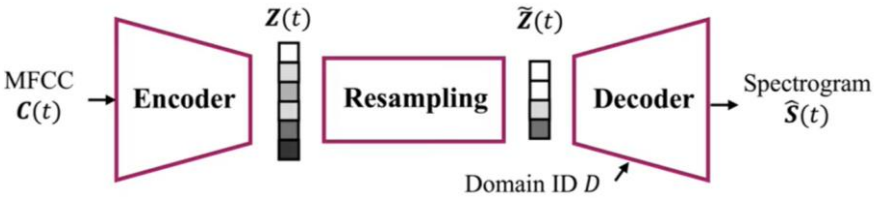
Student : Sian-Yi Chen

Advisor : Tay-Jyi Lin and Chingwei Yeh

Outline

- Action item
  - 找一篇關於 disentangled representation 有對應 open source 可以直接上手的相關論文研究

- Status report
  - 之前報告的論文 [1] 沒有提供 open source，因此找了另一篇 [2] 與 disentangled representation 相關且有 open source 的論文
  - [2] 論文概要
    - 在大多數非平行語音或是演算法中並不會轉換 prosody 這個重要資訊，作者提出了 AutoPST (Autoencoder-based Prosody Style Transfer)，其中架構由編解碼器夾著一個 resampling 所組成 (圖一)，在 resampling 這個 model 會藉由對時間重採樣來取得一個模糊的節奏 (obscure rhythm)，這樣解碼器就難以猜到原始語音的節奏資訊。
  - 實作進度
    - 開始訓練時發生程式錯誤，嘗試了一些方式但都沒有解決
    - 目前處理方式：
      1. 在 GitHub 尋求作者協助
      2. 向查理學長借 IBM AC922，測試是否是 GPU 的關係 (目前 IBM AC922 環境處理到一半，還有一些檔案與套件尚未處理完成)
      3. 還在理解程式錯誤訊息，並嘗試自行解決



圖一

[1] A. Polyak et al., “Speech resynthesis from discrete disentangled self-supervised representations,” arXiv:2104.00355 [cs.SD], Jul. 2021
[2] K. Qian et al., "Global rhythm style transfer without text transcriptions," arXiv:2106.08519 [eess.AS], 2021.

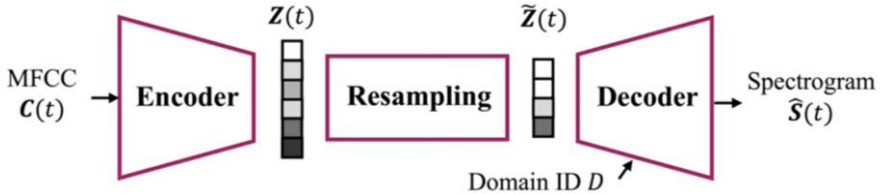
附錄

論文摘要

在大多數非平行語音或是演算法中並不會轉換 prosody 這個重要資訊，作者提出了 AutoPST (Autoencoder-based Prosody Style Transfer)，其中架構由編解碼器夾著一個 resampling 所組成 (圖一)，在 resampling 這個 model 會藉由對時間重採樣來取得一個模糊的節奏 (obscure rhythm)，這樣解碼器就難以猜到原始語音的節奏資訊，增加解碼器的難度已達到效果更好的去學習韻律特徵。

那什麼是節奏資訊呢？

假設我們講了一個單詞為 CAT，它可以拆分成 3 個音素，說話慢的人，它的這個單詞就會由較多的音素所構成，而說話快的人，所需的音素就較少。



圖一

作者在論文中表示，在觀察中發現人類說話速度改變時，音素縮放的比率是不均勻的，其中母音的變化率大於子音、輔音，因此 resampling 需要去模仿這樣的機制。

因此利用上、下採樣來模仿人類的行為，並訂定一些規則來設計 resampling model。

GitHub 專案所需環境、套件以及檔案

Dependencies

- Python 3.6 — 查看 python 版本 python –version **OR** python –V
- Numpy — conda install numpy
- Scipy — conda install scipy
- PyTorch == v1.6.0 — conda install pytorch=1.6 (python 3.6 版本只支援到1.4)
- librosa — conda install -c conda-forge librosa
- pysptk — pip install pysptk，需要安裝 VS C++
- soundfile — pip install SoundFile
- wavenet_vocoder `pip install wavenet_vocoder==0.1.1` for more information, please refer to https://github.com/r9y9/wavenet_vocoder — pip install wavenet_vocoder==0.1.1

To Run Demo

Download pre-trained models to `assets` — pretrained_models.zip - Google 雲端硬碟

To Train

Download training data to `assets`. The provided training data is very — vctk16-train-wav.zip - Google 雲端硬碟

Open chullin opened this issue yesterday · 0 comments

How to solve the problem that repeats has to be Long tensor

chullin commented yesterday

Sorry, I'm not familiar with English grammar, please forgive me if I offend.
I want to try to execute this Github project, but failed.
The only changed part of the program is (Because i don't have GPU)

(prepare_train_data.py)

> device = 'cuda:0' (Change to the following line)
> device = torch.device("cuda" if torch.cuda.is_available() else "cpu")

Run >python main_1.py
Problem Description : RuntimeError: repeats has to be Long tensor

```
Finished loading the 4 Utterances training dataset...
Start training...

Traceback (most recent call last):
  File "main_1.py", line 41, in <module>
    main(config)
  File "main_1.py", line 22, in main
    solver.train()
  File "D:\Work_Space\off_line_speech_converter\paper_speech_resynthesis\Global Rhythm Style Transfer Without Text Trans
criptions\AutoPST-main\solver_1.py", line 106, in train
    spk_emb)
  File "C:\Users\User\Anaconda3\envs\AutoPST\lib\site-packages\torch\nn\modules\module.py", line 532, in __call__
    result = self.forward(*input, **kwargs)
  File "D:\Work_Space\off_line_speech_converter\paper_speech_resynthesis\Global Rhythm Style Transfer Without Text Trans
criptions\AutoPST-main\model_autopst.py", line 167, in forward
    cd_short_sync = self.pad_sequences_rnn(cd_short, num_rep, len_spect)
  File "D:\Work_Space\off_line_speech_converter\paper_speech_resynthesis\Global Rhythm Style Transfer Without Text Trans
criptions\AutoPST-main\model_autopst.py", line 152, in pad_sequences_rnn
    code_sync = cd_short[i].repeat_interleave(num_rep[i], dim=0)
RuntimeError: repeats has to be Long tensor
```

Would i ask for help ?
I will be grateful for any help you can provide.

Assignees
No one assigned

Labels
None yet

Projects
None yet

Milestone
No milestone

Linked pull requests
Successfully merging a pull request may close this issue.
None yet

Notifications Customize

Unsubscribe

You're receiving notifications because you authored the thread.

1 participant