

數位語音專題 第二階段

2018/10/11

End-to-end Speech Recognition

- 助教：劉浩然 (r07922013@ntu.edu.tw)
- 第一階段大家已經實做了最經典的HMM-GMM/DNN ASR
- 在這個專題會進一步研究各式各樣Deep learning based的ASR
(e.g. Seq2seq, CTC, Neural Transducer, ...)
- 最主要的目的就是整合複數module, 達成End-to-end的ASR
- Roadmap
 - 認識最新的ASR技術 (Paper Reading)
 - 研究前人的不足或是缺點 (Study/Implementation)
 - 提出新的Idea/Solution (Publication, not necessary)
- Requests - Time & Enthusiasm

TTS(Text-to-Speech)

- 助教：劉達融(givebirthday@gmail.com)
- 瞭解現在tts技術，包括tacotron, wavenet等等
- 其實合成聲音不是難事，但是要合成出高品質、沒有機械音的聲音是困難的，而這次專題將著重在這點。
- 利用類神經網路聲碼器(Neural Vocoder)，例如Wavenet Vocoder、FFNet Vocoder達成高品質的合成語音。

語音及文字的嵌入表示與對齊

- 助教：陳奕禎 (r06942069@ntu.edu.tw)
- 目標：語音及文字的嵌入表示之共同特徵與對齊
- Dataset：英文朗讀的聲音及文字，其中聲音與文字並沒有事先對齊 (align)
- 利用自動編碼器及生成對抗網路的技術，分別學習語音及文字的嵌入表示，並利用其共同特徵來對齊

文體風格轉換 Text Style Transfer

- 助教：陳冠宇 (r06942070@ntu.edu.tw)
- 讓機器自動產生句子是機器學習(machine learning)中一個很常見的應用。而在這個題目中我們希望可以控制機器產生的句子，使其具有我們想要的特性(style)。
- 這個專題中會先學習現在常見的文句生成模型(RNN, seq2seq)，再對這些模型做修改來達到目的。

Video Caption Generation and Evaluation

- 助教：莊舜博 (f04942141@ntu.edu.tw)
- 從文章總結出大意(Summary)或者從圖像中辨識出物件(Image Recognition) 為機器學習中常見的應用，藉由對輸入的分析與學習使得機器得以達到理解，並產生相對應的輸出。
- 在此專題將會以影片為主軸，輸入影像與聲音，讓電腦輸出相對應的文字敘述。專題會著重在聲音對文字生成的影響，而如何衡量輸出敘述的正確性也為課題之一。
- 此外，如何使用Audio-to-Image Generation的方式提升Video Caption 的表現也為研究方向之一。

Deep Reinforcement Learning

- 助教：茅耀文 (r06921053@ntu.edu.tw)
- Deep Reinforcement Learning (Deep RL, DRL)
- RL是一種問題的設定,行動者(Agent)要在環境中做一連串的動作(Action),而其目標是最大化利益(Reward)總和。
- DRL即是用深層學習的技術來處理強化學習(Reinforcement Learning)的問題。
- 本專題會學習RL的各種數學原理和現行知名的DRL方法。在此之上,將這些方法應用到各種遊戲上面,並研究現行方法的不足之處和可能的改進方式。

聲音轉換

Voice Conversion

- 助教：葉政杰 (r06942067@ntu.edu.tw)
- 一段語音的訊號包含了許多複雜的資訊，例如：語言、說話方式、情緒、語者、環境等。若能理解不同資訊造成的訊號變化甚至以生成模型來生成自然語音(Natural Speech)，則可以有非常多的應用。
- 本專題目前主要專注在韻律可變(Rhythm-flexible)、不需平行資料(Non-parallel)的聲音轉換，以及類神經網路聲碼器(Neural Vocoder)的訓練。將利用到深度學習中VAEs以及GANs的技術。

使用深層學習之語音資訊摘要

Automatic Speech Information Summarization with Deep Learning

- 助教: 王耀賢 (r06944019@ntu.edu.tw)
- 自動摘要(Automatic Summarization)可以用機器自動找到
- 一段語音或文字的重點,將其轉化為比較簡潔且能大致描述整段資料內容的形式。在資訊爆炸的今天,自動摘要可以幫助我們節省許多瀏覽資料的時間,尤其是對於不易瀏覽的語音或多媒體資訊;在資訊搜索上,自動摘要可以用以篩選、排序、呈現搜尋結果;此外,行動裝置的流行也帶來另一個重要應用:因為行動裝置螢幕大小受限,自動摘要可幫助更佳地呈現資料在行動裝置上。
- 在這個專題中,我們會學習到各類自動摘要的基本原理及架構,並研究如何將深層學習(deep learning)應用到概念式自動摘要(abstractive summarization)上。

Question Answering

- 助教：李佳軒 (r06944037@ntu.edu.tw)

1. END-TO-END spoken question answering

過去的spoken question answering都是兩階段，ASR系統再接QA系統，可以參考這篇paper

(https://www.isca-speech.org/archive/Interspeech_2018/pdfs/1714.pdf)

- 實驗室已經有數個語音corpus (Interspeech, SLT)，會希望加入的學生可以在這些corpus上嘗試不同model，提出新想法

• 2. SQuAD2.0 :

ACL best short paper propose了SQuAD新版本Data，再過去只要能夠登上leader board就是ICLR, ACL, EMNLP的保證，現在各家MODEL離human performance還相差甚遠。

題目登記方式

- 寄信到 r07942076@ntu.edu.tw
- 標題:[數位語音專題] 姓名1 姓名2
- 內容包含:
 - 兩個組員姓名
 - 題目志願序5個
- 如志願序衝突, 則先寄者優先