

數位語音專題 第二階段

2019/03/21

文體風格轉換 Text Style Transfer

- 助教: 李仲翊(r07942080@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。

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、FFTNet Vocoder達成高品質的合成語音。

Video Caption Generation and Evaluation

- 助教: 莊舜博 (alex82528@hotmail.com.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。

聲音轉換

Voice Conversion

- 助教：許博竣(r07942095@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)是自然語言處理的經典問題之一, 希望讓機器自動抽取文章大意並且產生摘要。
- 專題開始會先給大家念些相關的文章, 讓大家了解目前主流的模型。再來會希望大家能夠實作基礎的attn-seq2seq model, 以及進階的model像是pointer network或者google transformer。
- 最後會希望大家能改善現有model的表現, 可能的方向有利用其他輔助的task或者大量未標註的文章來幫助機器理解更深層的文意。

Question Answering Track

- 助教：李佳軒 Michael (r06944037@ntu.edu.tw)
- QA這個TASK用意是讓機器理解文章以及問題背後的語意。模型中最为關鍵的機制是Attention，過去的attention是input sequence跟output sequence的ALIGNMENT PROBLEM；但過去一年內self-attention被應用在NLP各領域中，並得到多項STATE-OF-THE-ART的表現。關於SELF-ATTENTION的設計哲學再本質上必須要跟過去的ATTENTION不同，如何改良SELF-ATTENTION會是我們這個專題的主題。
- 另一個可能的目標是改良BERT。BERT是目前最大型的language model，被發現對於improve downstream task非常有用。然而目前在QA task上，如何將BERT與QA model結合還未有好的設計，我們也可以選擇做這個題目。

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, 從今日專題下課起開始計算。