

# 第二階段專題

# 影片敘述之自動生成

## Video caption generation

- ▶ 助教: 莊舜博 ([alex82528@hotmail.com.tw](mailto:alex82528@hotmail.com.tw))
- ▶ 從文章總結出文章大意(summary)或者是從照片辨識出其中物件(image recognition)為機器學習(machine learning)中常見的應用，都是教電腦學習我們輸入的資料，並基於機器的理解產生相對應的輸出。
- ▶ 在這個專題中將會以影片為主軸，給機器看一段影片及其聲音後，使其學習理解影片內容，並自動產生與影片相關的敘述。

# 文體風格轉換 Text Style Transfer

- ▶ 助教：陳冠宇 ([r06942070@ntu.edu.tw](mailto: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。

# 對話生成

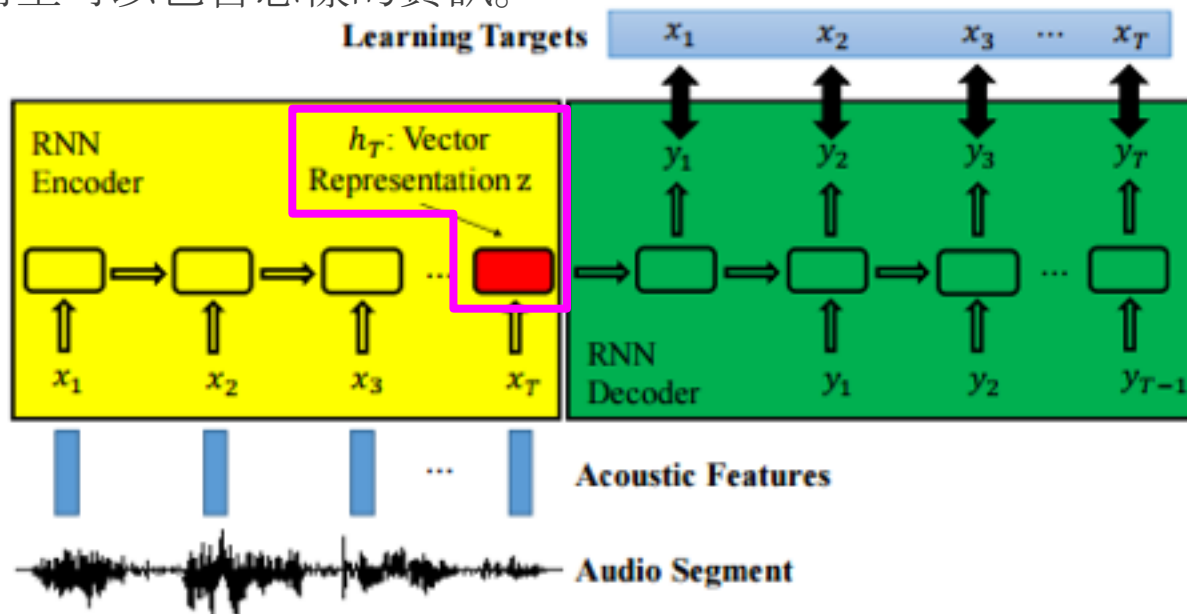
## Dialog generation

- ▶ 助教: 吳元魁 ( [b02901137@ntu.edu.tw](mailto:b02901137@ntu.edu.tw) )
- ▶ 根據輸入的句子輸出相對應的對話，助教自己是用 `pytorch` 來做 `implementation`，模型從比較基本的 `sequence-to-sequence` 和 `attention mechanism`，到比較新的 `generative adversarial network`，還可以嘗試把感情放入對話的生成。
- ▶ 要解決的問題包括
  - ▶ `generic response`(輸出一些通用的回答， `ex: I don't know.`)
  - ▶ `gradient` 在 `gan` 裡面無法做 `backpropogation`
  - ▶ 生成句子如何合理的給予 `reward`
  - ▶ 如何產生特定感情的句子

# 聲音訊號的向量表示法

## Audio word2vec - discovering phonetic structure from audio

- ▶ 助教: 陳奕禎 (r06942069@ntu.edu.tw)
- ▶ 語音版的Word2Vec, 將聲音直接轉為向量, 並探討這樣的向量可以包含怎樣的資訊。



- ▶ 這是最基本的RNN架構, 然後接著呢?

# 聲音訊號的帶語意的向量表示法

## Audio word2vec - discovering semantic from audio

- ▶ 助教: 黃邦齊 ([zyi870701@gmail.com](mailto:zyi870701@gmail.com))
- ▶ 小孩子學說話時，看到爸爸時知道要喊“BaBa”，看到媽媽時要喊“MaMa”，但其實都還不知道“爸爸”和“媽媽”這兩個詞分別是甚麼意思，而是久在生活中慢慢學會。
- ▶ 本專題是希望機器在聽了大量聲音及讀了大量資料之後可以學到直接由聲音訊號發現語意，並把學到的語意呈現在問題表示法中。

# 使用深層生成模型之語音轉換

## Speech Transformation with Deep Generative Model

- ▶ 助教: 葉政杰 ([r06942067@ntu.edu.tw](mailto:r06942067@ntu.edu.tw))
- ▶ 語音轉換(Speech Transformation)是當今相當重要的課題。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。
- ▶ 本專題將會使用現今深度學習中最火紅的生成對抗網路(Generative Adversarial Networks, GANs)的各種不同衍生架構，如Cycle-GANs，來實作語音轉換課題，並主要專注在非母語與母語之間的轉換(non-native-to-native)。

# 通用問答系統

## Real QA

- ▶ 助教：李佳軒 ([chiahsuan.li@gmail.com](mailto:chiahsuan.li@gmail.com))
- ▶ 目的是讓機器擁有看完文章後有回答問題的能力，實驗室學長們已經實作了兩種在托福聽力測驗(input為語音辨識過的文本)上達到40%以上正確率的模型。這個研究領域已經破解了如何從文章中找出特定單字當作答案的dataset。目前正在攻略讓machine自己挑選長度不定的字詞片段。
- ▶ 這個題目不需要特別的NLP知識，如果要作需先學習memory network的基本知識。未來的方向可能包括：
  - (1) 嘗試不同模型去衝高托福測驗的正確率
  - (2) 將托福聽力測驗的聲音訊號加入現有模型進行訓練
  - (3) 讓機器擁有基礎知識，不必閱讀文章即可回答問題

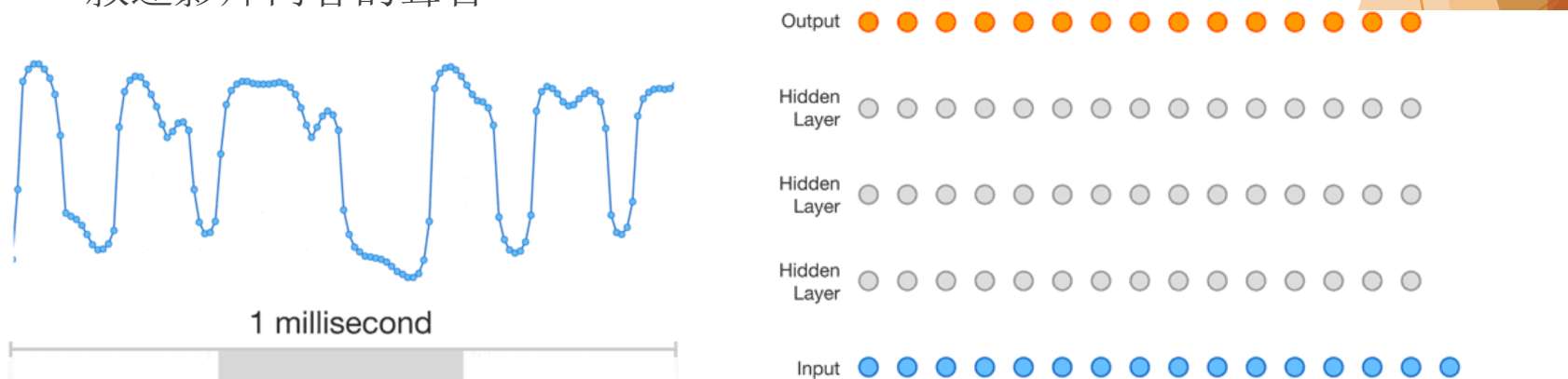


# Deep Reinforcement Learning

- ▶ 助教：茅耀文 ([r06921053@g.ntu.edu.tw](mailto:r06921053@g.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动作 (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。

# Wavenet: A Generative Model for Raw Audio

- ▶ 助教：楊棋宇
- ▶ **Wavenet**是一個可以生成原始語音的深度生成模型，此模型可以生成類似真人發音的語音，並且品質比任何現有的模型所生成的語音都要好
- ▶ 在此專題我們會利用烹飪影片中的聲音來訓練模型，讓模型可以生成類似於烹飪影片中的人聲，更進一步會加入聲音特徵作為額外的輸入，希望能控制wavenet模型生成出敘述影片內容的聲音



# 以增強學習解決檢索問題

## DRL for retrieval

- ▶ 助教：鍾佩宏([chung95191@gmail.com](mailto:chung95191@gmail.com))
- ▶ 近年來網路上多媒體數位內容(Multimedia Content)大幅增加，語音數位內容之檢索(Spoken Content Retrieval)也因而大受重視。
- ▶ 本專題之目標放在互動式檢索(Interactive Retrieval)，採用增強學習(Reinforcement Learning)的演算法來學習出最佳系統決策，並研究各種深層強化學習中，找出能改善互動式檢索的方法，解決語音內容檢索的問題。

# 使用深層學習之語音資訊摘要

## Automatic Speech Information Summarization with Deep Learning

- ▶ 助教：盧柏儒(r05942070@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。

# 語音上的攻擊樣本

- ▶ 助教：吳思霖(b01901118@ntu.edu.tw)
- ▶ 這是語音版的**Adversarial example**，理想的攻擊樣本人類聽起來會與原本聲音無異，但語音辨識系統卻會辨識出迥異的結果。
- ▶ 可以先從圖片上的攻擊樣本入門，了解架構後轉而實作語音的版本，方向大致有（1）**White-Box**、（2）**Black-Box**、（3）**Targeted Attack**，如果同學有興趣也可以轉而做防禦攻擊樣本的辨識系統。

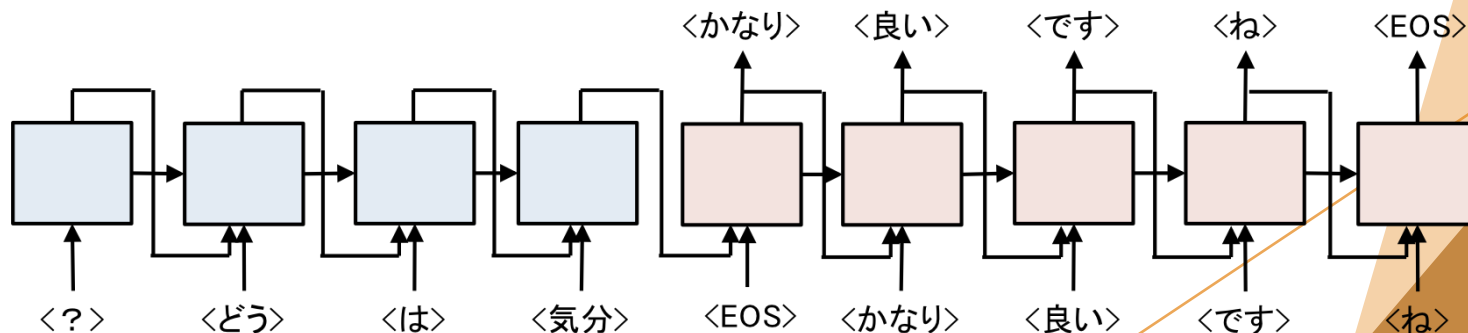
# 生成對抗網路在自然語言處理上的應用

## GAN on Text Processing

- ▶ 助教：段逸林 (b02901048@ntu.edu.tw)
- ▶ 現今生成對抗網路 (GAN; Generative Adversarial Network) 在影像上的應用十分廣泛且成功，相較來說，在處理自然語言時卻沒有如此順利，首先會遇到的問題便是自然語言本身離散的特性。因此近一年才開始有許多研究集中在處理將生成對抗網路應用到自然語言處理相關的題目(例如：text generation, chatbot, style transfer, 等等)。
- ▶ 在這專題中，我們會先學習各種GAN的原理以及在這類題目中常用的seq2seq model，並討論要處理的text processing相關的題目，研究GAN在目標題目的應用和進步。

# Sequence to sequence learning and its application on Speech

- ▶ 助教：劉達融([b01901079@ntu.edu.tw](mailto:b01901079@ntu.edu.tw))
- ▶ Sequence-to-sequence learning 在語音上有很多應用，因為一句話、一個語音訊號等都可以用一串向量表示
- ▶ 相關題目：語音辨識、機器翻譯、文字生成、對話機器人



# 深度學習流言終結者

- ▶ 助教：黃淞楓
- ▶ 目的是要研究長期以來被視為黑盒子的 NN 內部架構是否能夠辨識、記憶 data distribution，並針對 data distribution 去架構出相對應的 function。
- ▶ 透過給定已知 distribution 的 data，例如二維的 mixture of two Gaussians，比較現有的 DNN-based recognizers 的辨識效果，並且針對 model 的架構去做探討，如中間每層 neurons 數值的變化。
- ▶ 研究 data 的 dimensionality 與 data size 之間的比例 (data-size-to-dimensionality ratio) 對 model 的影響。假設我們預期近期的 DNN 方法在 low data-size-to-dimensionality ratio 的時候可以有好的表現，那我們可以先找個無法處理這種狀況的傳統統計學方法去實驗，顯示傳統方法的不足，然後再用近期的 DNN techniques 去重複1.的實驗，並且彼此做比較，找出各 model 的優缺點與 DNN 強力的原因。



# 題目登記方式

- 寄信到 [r06942067@ntu.edu.tw](mailto:r06942067@ntu.edu.tw)
- 標題：[語音專題] 姓名1 姓名2
- 內容包含
  - 兩個組員的姓名
  - 題目志願序
- 如志願序衝突，則先寄者優先