

「科技大擂台 與AI對話」

熱身賽初賽

講者：李宏毅



「科技大擂台 與AI對話」

2017

熱身賽

初賽

8/1 - 9/1

半決賽

9月中

決賽

10月中

2018

正式賽

初賽

3月－6月

- 讓大家在拿高額獎金前先暖個身
- 參加對象：大專院校、高中職學生 (一到四人一組參加)
- 首獎有十萬
- 有高額獎金
- 但跟這段影片沒什麼關係
- 全民皆可參加

熱身賽初賽－試題

以下出題模式
僅限熱身賽

- 機器在聽一段人類對話後，預測人類的下一句話

初賽直接提供字幕檔

A: 柏翰謝謝你

B: 不客氣

A: 不好意思耽誤你下
課的休息時間

選項：

(1) B: 我很喜歡下課時間

(2) B: 都這麼熟了不用說
客套話

.....

- 試題下載
 - 範例試題 (**development set**)：8/02 後可以在科技部官
網上下載 50 題的範例試題和其答案
 - 測試試題 (**testing set**)：8/07 釋出

熱身賽初賽－試題

以下出題模式
僅限熱身賽

- 機器在聽一段人類對話後，預測人類的下一句話

初賽直接提供字幕檔

A: 柏翰謝謝你

B: 不客氣

A: 不好意思耽誤你下
課的休息時間

選項：

(1) B: 我很喜歡下課時間

(2) B: 都這麼熟了不用說
客套話

.....

熱身賽初賽不需要做語音辨識，
但半決賽和決賽仍然需要語音辨識

熱身賽初賽 – 作答方式

- 可以使用任何程式語言、套件、深度學習框架
- 可以用任何技術讓機器選出正確答案，除了 ...
 - 人自己去找出正確答案
- 請不要在初賽嘗試任何作弊的方式
 - 你可以輕易地用作弊的方式在試題上得到 100% 正確率
 - 參加半決賽的隊伍要繳交程式碼和模型供主辦單位檢查，繳交方式、格式另行公告
 - 半決賽的題目將來自你無法取得的資源，且為現場測試
 - 如果你用作弊的方式進入半決賽，但沒有真正的技術，我會在半決賽大力嘲笑你

熱身賽初賽－官方提供資源

- 運算資源：申請辦法詳見官網
- 訓練資料：公視節目字幕檔
 - 請自行思考要如何利用這些字幕檔
 - 因為初賽所有的題目都出自公視節目，所以公視的字幕檔對結果應該會很有幫助
 - 當然，題目來源的那幾集節目的字幕檔並不會包含在官方的訓練資料中 😊
 - 從試題中推論出我們使用哪些節目的哪幾集出題，請不要自行在網路上找出那幾集節目加入訓練資料
 - 不然你會在初賽拿到很好的成績，但半決賽很慘，然後被大家恥笑
 - 比賽中可以使用自行蒐集的資料來訓練你的機器
 - 也歡迎在討論區分享你找到的資源

熱身賽初賽 – 繳交答案

- 本競賽透過 Kaggle 平台繳交答案
 - 有關 Kaggle 的相關知識和使用方式請見官網
- 在比賽期間 (8/01 – 9/01) 每組每日可以上傳五次答案
- 初賽將於 9/01 23:59 結束

如何開始呢？

- 先熟悉深度學習的各種技術
 - 官網上有深度學習相關教學資源，以及機器學習教學資源
- 學習一個自己喜歡的深度學習框架
 - 官網上有各種深度學習框架教學文件
- 有任何深度學習相關問題歡迎在討論區上討論
 - 競賽期間，討論區上所有時段都會有熟悉深度學習技術的小編駐板
 - 小編ID: AI mentor

答題方向提示

- Sequence-to-sequence model
 - Input: a sequence, output: a sequence

A: 柏翰謝謝你

B: 不客氣

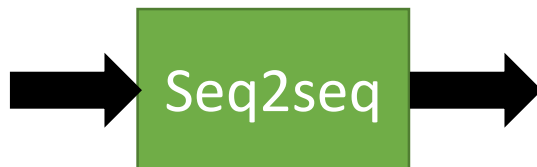
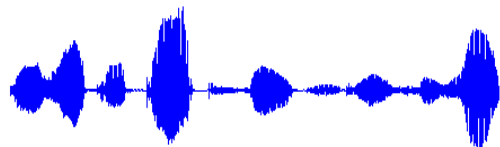
A: 不好意思耽誤你
下課的休息時間



不用說客套話啦
(選一個最接近
的選項)

你可以在官網的教學區找到
Sequence-to-sequence model 的相關教學

Sequence-to-sequence model



“大家好，歡迎大家來修機器學習”

Speech Recognition

“機器學習”



“Machine learning”

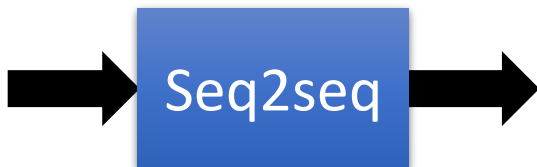
Machine Translation



“A young girl is dancing.”

Caption Generation

“Hello”

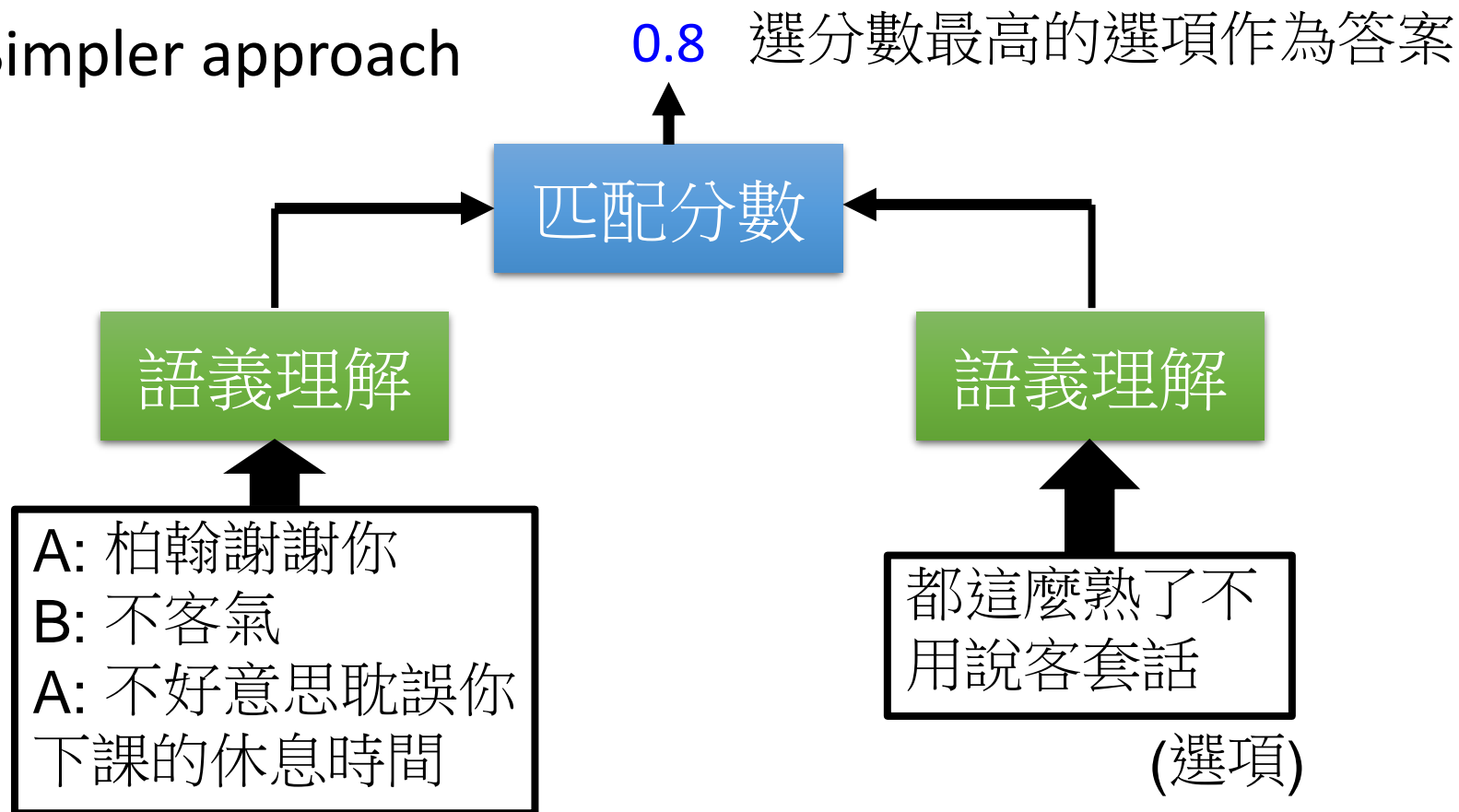


“Hello. Nice to see you.”

Chat-bot

答題方向提示

- Simpler approach



Word Embedding 相關技術你可以在官網的教學區找到

初賽完整攻略

報名前

- 組隊，1~4 人一隊，本競賽所有活動都以組為單位
- 隊伍中一人(隊長)註冊 Kaggle 帳號

8/01 報名開始

- 隊長在官網報名，報名後在官網上取得官方訓練資料和範例試題

8/07 試題釋出

- 機器在正式試題上做答後，就可以開始上傳結果

8/15 報名截止

- 要報名就快，不要猶豫了

9/01 初賽結束

- 接下來還有半決賽和決賽喔