Ai演講

S10767004吳瑋哲

「鑑往知來:子非人，安知人知語?」講座談論了AI在自然語言處理的核心技術，如何降低資訊傳遞上的障礙，一直是人工智慧的典型應用範例。演講由文本分析、問答系統、機器翻譯等應用出發，說明電腦非人，但是能理解人類語言的關鍵。提到了自然語言處理的應用與挑戰。

第一部分:機器翻譯

機器在不同文本間的翻譯最早就從冷戰時期的情報收集開始了，近年生活也應用到了導覽、圖像與語音整合等翻譯。

第二部分:問答系統

分析問題找出問什麼，分析內容進行文件探勘並擷取正確答案，最後再透過自然語言處理、機器學習等來選擇答案，達到標的追蹤、輿情分析等目的。當然在自然語言處理方面有一些典型的問題，如岐義解析中的詞彙層次、語法層次、語義層次、容錯力、強健性。如何理解語言單位值得探討。且要分類語詞就要掌握共通性，從詞性、語義、句法、相依、言談、意見、情感、立場等類別等方面著手。分佈式表示(Distribution Representation)是一個重點，要了解詞彙的意思，關鍵是伴隨出現的詞彙，如何在相似甚至相同的語境界定上下文的範圍，值得研究，也可以利用向量計算來計算詞彙關係的緊密度，而維度太高導致太稀疏的缺點，可以透過降維處理來解決。至於分散式表示(Distributed Representation)是向量空間中詞彙的表示，透過類比、向量、相似關係等計算，以捲積神經網路產生句子向量。問答系統應用方面，在問題→文件檢索→答案選擇步驟中，涉及到相依剖析、路徑匹配、句子相似計算等等。Ai問答系統已應用於google多語言神經機器翻譯系統、性能評估等等。

第三部分:意見探勘應用

這部分舉了許多例子，如旅館評論中我們可以分析出外顯、內隱意見，也可辨別詞彙的正或負面性;領域相關詞彙可做極性分析;StockTwits對大量使用者的發文、標記、表情資料進行分析;財務和經濟資料中從資料來源、寫作風格追朔到官方文件、新聞、社群媒體等等。市場情緒和發文者情緒是不同的，因為取決賺錢與否。

第四部份:總結

最後演講進行討論，得出領域知識對於自然語言處理很重要、領域的轉移對於系統效能有很大的影響、資料與知識的整合是重要趨勢、語言理解與世界知識、常識等外部知識也相關、建立人類常識、因果關係、五感資訊等的多模態知識庫為發展重點等結論。整場演講言簡意賅、詞彙達意，是一場很棒的講座。

第五部份:問答

問答階段，學生提出了兩個詞彙如何在向量空間中定義、天氣觀測與表情辯識如何進行自然語言處理等問題，問的都很有水準。印象最深刻的是其中一人提出的問題-神經網路使用在轉換資訊時，我們對過程的複雜權重並不了解，通常我們只知道輸入與輸出但關於過程中它的理解我們卻沒全然了解，擔心未來有AI學壞的情形。講者也以分析ALPHA GO如何決定下棋步驟來精進人類下棋招數為例子，表示目前要解開這「黑盒子」非常困難，是未來的研究的趨勢。我自己也覺得這個問題問的很精銳。