人工智慧心得0602

探索20-4講座：鑒往知來：子非人，安知人之語？—談自然語言處理 / 陳信希教授

數位三 S10755040 王綉孋

透過這個演講讓我了解到有關語言處理的基本觀念以及相關應用，主題是「論ㄌㄨㄣˋ語: 子非人，安知人之語？」，也就是討論語言這個深奧的問題，電腦是機器不是人，是否能擁有人的智慧呢?語言是人類溝通重要的工具，包括聽說讀寫等，電腦能否聽、說、讀、寫、理解人類的語言，就是自然語言重要的關鍵。講者提出一些相關應用讓我們更加了解，機器翻譯、問答系統、意見探勘三大應用。

最經典的語言處理應用是機器翻譯，如在不同語言間的文本自動翻譯原語言和目標語言、使用詞彙和結構差異、原始語言文句分析與目標語言文句生成，由早年冷戰時期的情報收集、到近年生活面向的各種需求，情報傳送像是加密與解密，將語言翻譯成人類看不懂的編碼，藉由金鑰解碼，到現在國際化使用各種不同語言並需要翻譯，與圖像文字或語音整合翻譯與地理位置整合的方言翻譯等。

分析問題最重要是要找出在問甚麼，先分析內容擷取正確答案，計算支持或反駁資訊的信心度等，其延伸為文本分析，文本分析，對新聞、科技論文、電子郵件、網頁、部落格貼文、微網誌、病例等不同類型來源的數位資料，進行文件探勘，可以應用於市場產品資訊、政治意見追蹤、社群網路分析、熱門議題分析。但自然語言處理還是有一些典型問題，像是代名詞問題，代名詞在此指涉的是什麼，語法錯誤、同義詞、歧異性、未知詞(沒有被收錄的詞彙)等，從詞彙面向來看一字多義該如何翻，在結構面向，英文與中文語句不同，介係詞片語位置中英文位置不同等。

在語言上的規律性，因有規律性，就可以用電腦來分析，在分析當中很重要的事情就是分類，情感類別、立場類別、語意類別，可以用詞向量分析、上下文分析，要真正讓機器能夠擁有如同人類語言能力，機器翻譯是否已經達到人類翻譯的境界，語言是不斷在改變的，會有新的詞彙出現，因此，機器還是需要經過層層處理，才能夠更接近「自然」語言。

<NOTE>

語言單位、成分、或單元

詞素(Morpheme): -ly, -ful, -able, un-, 們

詞彙(Word): bat, cat, 台灣大學

字元(Character); a-z, A-Z, 台灣, 大, 學

n-連詞(N-gram):台灣,灣大,大學,台灣大,灣大學 多詞表達(Multiple Word Expression) 網絡紅人,網紅

命名實體(Named Entity): 台灣積體電路股份有限公司,台積電,王建民

子句 (Clause)

句子(Sentence)

段落(Paragraph)

段落(Passage)

文件 (Document)

多文件(Multi-Document)



















