演講摘要摘要/重點筆記 材料三 s10767022 張沛瑄

如同演講陳述的目前「語言文字」是人與人之間的溝通的主要工具，而 AI 人工智慧是目前的趨勢，必定會使用到電腦，那電腦是否能理解聽、說、讀、寫，為是否擁有人的智慧的重要指標。

從背景切入應用和挑戰主要分為**三類**：

1.機器翻譯

主要是透過機器來把英文轉換成中文，例如生活方面：出國看不懂外國的文字，可以使用英(日)中翻譯。而翻譯技術會隨著時間增長，而有所進步，演講者透過十年觀察，輸入中文翻譯成英文，看有什麼差異及進步。結果可以發現在翻譯結果上有很大的突破。

可以分成兩個面向來看：

- 詞彙面向：同一個英文詞彙翻譯會隨年代不同來做修正。
- 結構面向：介系詞片語的位置中英文不同。

2.問答系統

- 分析問題：找出問題問什麼
- 分析內容：擷取正確答案
- 計算支持或反駁資訊的信心度
- 自然語言處理、資訊檢索、機器學習、知識表示和推理及大規模平行計算等。

例如：問題是什麼➔文件檢索（可能含有答案的資訊）➔答案選擇（但有代名詞指涉問題：主詞是他（不是資訊的答案）還是諾貝爾獎人名等）

3.意見探勘

- 像市場產品的資訊或社群網路的分析

而自然語言也存在著一些常見的問題，例如：

- 歧異解析(詞彙層次、語法層次、語義層次)
- 容錯力：字打錯或語法錯誤
- 強健性：領域改變、網路符號及表情符號

簡單的自然語言處理

1. 概念表示：符號表示法（不同語言用不同方式呈現概念），但是其符號本身會帶有歧異性。所以要了解符號到底代表甚麼意涵。
2. 語言單位、成分或單元：詞彙、字元、n-連詞、多詞表達...等。
3. 分類：掌握共同性(把所有概念整合)

- ✧ 詞性類別：利用分詞及斷詞。
- ✧ 語義類別：例如：夕陽，斜陽，落日，都是同一種詞彙(類別)意思。
- ✧ 句法類別：使用人給的符號，而電腦掌握規律性，所以要知道句法的類別。

- ✧ 相依類別：如何找出詞與詞的關聯性。
- ✧ 言談類別：時序(temporal)-因果(contingency)-轉折(comparison)-推展(expansion)。
- ✧ 意見類別：有正面和負面針對某個議題。
- ✧ 情感類別：憤怒，開心...。
- ✧ 立場類別：贊成或反對對某個議題。

符號計算：

詞彙和類別都是以符號呈現，電腦做匹配的動作，而匹配是基本的符號計算

像夕陽與落日匹配失敗，所以要藉由外部資源（像同義詞）。由於符號計算無法計算詞彙關聯程度(像及物動詞/不及物動詞的概念)。

所以有

(1) 分佈式表示（distributional representation)：意思的產生來自使用，要了解詞彙的意思，關鍵是伴隨出現的詞彙。

分佈式假設：如果兩個詞的上下文相似，則這兩個詞的詞義是相似的。

例如：語境（上下文）像關門，有把門關上，打烊，停業的意思，但跟據上下文的不同會有不同的意思存在。可以用向量的方式表示關門。每個詞彙都是以高維度向量表示（夾角越小，詞彙關係越緊密）

優點：可計算語義關連程度。缺點：維度太高，太稀疏，所以要降維處理。

(2) 分散式表示（distributed representation)

①. CBOW：用上下文的詞來預測目標詞

②. Skip-gram：以當前的詞來預測上下文的詞

而將詞彙轉成低維度稠密向量（word2vec）語法關係與語義關係叫類比運算。

word2vec 延伸：由符號的離散資料出現，改變成連續資料呈現。可用下列三項表示：

- 語言成分表示：詞向量、句子向量、段落向量等。
- 成分類別表示：詞性向量、詞義向量、句法類別向量。
- 用戶資料表示：用戶向量。

問答系統應用：相依剖析和路徑匹配（句子相似性計算：問答系統答案選擇）

意見探勘應用：電腦要知道像評論意思的真實意涵。