**自然語言處理演講重點筆記**

**電腦是否擁有人的智慧?**

語言文字是人與人之間溝通的主要工具

電腦能否聽、說、讀、寫、理解人的語言，為是否擁有人的智慧的重要指標。

**背景:應用和挑戰**

1.機器翻譯:不同語言間的文本自動翻譯

2.問答系統:分析問題及內容，找出在問什麼、擷取正確答案

3.意見探勘:對不同類型來源的數位資料進行分析

**自然語言處理的典型問題**

1.歧義解析:辭彙層次、語法層次、語義層次

2.容錯力:打錯字、語法錯誤、未知詞

3.強健性:領域改變、網路語言、表情符號

**自然語言處理**

概念表示

歧義

語言單位、成分、單元:詞素、字元、詞彙…等

**分類:掌握共通性**

詞性類別 語義類別 句法類別 相依類別 言談類別 意見類別 情感類別 立場類別

**分佈式表示**

意思 語境 上下文範圍界定 共同出現 向量計算

**分散式表示**

向量空間的詞彙表示 將詞彙轉換成低維度稠密向量 類比運算 向量運算 相似關係計算 以卷積神經網路產生句子向量

**問答系統應用**

相依剖析和路徑分析 句子相似性計算 問答系統答案選擇

**機械翻譯應用**

Google多語言神經機器翻譯系統 效能評估

**意見探勘應用**

旅館評論 領域相關詞彙極性分析 情感分類 財務和經濟資料

**討論**

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

資源在以機器學習/深度學習為基礎的自然語言處理扮演重要角色

稀少資源或有限資源機器學習/深度學習的研究，資料和知識的整合是重要趨勢

語言陳述可能含主觀、誇大、不實、諷刺…等資訊

語言理解不僅和語言有關，還可能需要用到外部知識