演講題目:子非人，安知人知語?

一.自然語言的應用及挑戰

1. 機器翻譯:不同語言間的文本自動翻譯

挑戰:原始語言和目標語言使用詞彙上的差異

應用:與圖像文字或語音整合的翻譯

1. 問答系統:問及回答問題

挑戰:分析問題，找出是在問甚麼

分析內容，擷取正確答案

技術:自然語言處理、資訊檢索、機器學習、知識表示與推理，大規模平行計算

1. 意見探勘: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、病歷等不同類型來源的資料進行文本探勘

應用:找旅館、市場分析、選情分析

二.自然語言一些常見的問題

1.容錯力

(1)打錯字

(2)語法錯誤

(3)未知詞

2.強健性

(1)領語改變

(2)網路語言

(3)表情符號

3.歧義分析

(1)詞彙

(2)語法

(3)語義

三. 淺談自然語言處理

1.概念表示:概念用符號表示，不同語言用不同符號表示，但符號有歧異性

2.語言單位、成分和單元:真實世界的概念以符號為基本單位組成句子或段落

四.掌握共通性:分類[自然語言處理]

1.詞性類別[詞性標記]:名詞、動詞、副詞……

2.語義類別[語義關係標記]:同義詞

3.句法類別[句法分析]:名詞片語、介系詞片語

4.相依類別[相依分析]:詞彙與詞彙的分類

5.言談類別[言談分析]:時序、因果、轉折、退展…..

6.意見類別[意見分析]:正面負面的詞彙

7.情感類別[情感分析]:憤怒、厭惡、恐懼、喜歡、悲傷、驚訝……

8.立場類別[立場偵測]:贊成或反對某一議題

五.分佈式表示

1.意思: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

2.分佈式假設:如果兩個詞的上下文相似，則這兩個詞的詞義相似

3.向量計算

(1)向量空間的詞彙表示

(2)將辭彙轉換成低維度稠密向量

(3)類比運算

(4)向量運算