

應用深度學習於自然語言之文句相似度分析

學生：林沛澤

指導教授：傅家啟

柯貞羽

邱國龍

王承佳

國立雲林科技大學工業工程與管理系

摘要

近年來人工智慧(Artificial Intelligence, AI)發展迅速，許多工作紛紛被電腦取代，客服系統也出現由線上系統問答取代過往專人接聽的趨勢，在人力與 AI 人工智慧更迭的同時，仍然有技術層面上的問題需要透過改善進而達到完善，例如：消費者常在輸入關鍵字後無法取得最佳解、由於資料庫缺少相關解答導致 AI 人工智慧搜尋不到相關問題的解答... 等問題產生，使得消費者在使用此系統後依舊無法解決問題。

在目前的機器學習領域中，自然語言處理 (Natural Language Processing, NLP) 是非常重要的應用之一，其中 Word2Vec 模型扮演極為重要的角色。為完善問答系統之設置與資料庫之健全，本組使用 Jieba(中文斷詞工具)達到斷詞準確，採用 Word2Vec 作為詞向量嵌入模型將文字轉變成向量，並輸入本組問答資料與 Wiki 資料庫來訓練詞向量，一共得到三組詞向量分別為 CBOW (Continuous Bag Of Words)、Skip-gram 與 Wiki 詞向量。得到模型後本組將測試資料代入模型中與訓練資料進行相似度計算，分別採用了 Cosine 餘弦相似度與深度學習進行計算。為了克服迴歸神經網路(Recurrent Neural Network, RNN)的梯度消失與梯度爆炸問題，我們以長短期記憶模型(Long Short-Term Memory, LSTM)與門控循環單元模型(Gated Recurrent Unit, GRU)分別建立基礎公文問答系統來計算問題與問題之間的相似度，並選用相似度最高的問題的答案來做為回應，且計算出各方法的準確率後進行績效評估與比較各個訓練方式間的差異性。