

利用自然語言處理進行自動新聞分類之研究

19

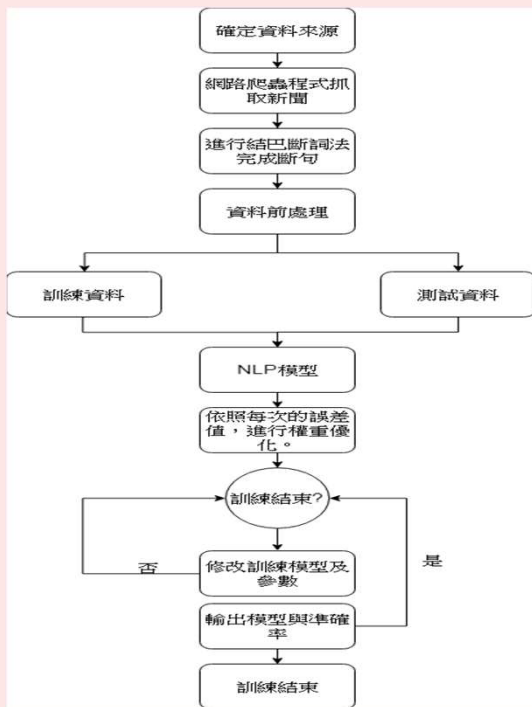
指導老師：李佳衛老師

專題學生：吳孟瑾、傅詞源

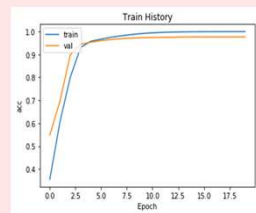
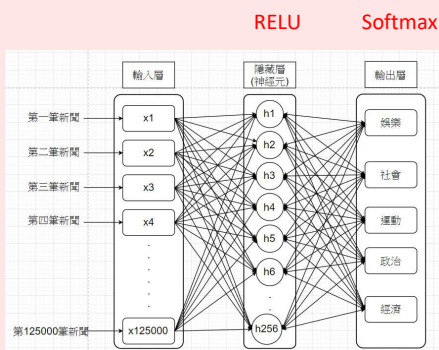
摘要

近年來，由於人工智慧(Artificial Intelligence, AI)崛起，使得機器具有更好的判斷能力，甚至優於人類。本專題是利用人工智慧的方法訓練機器，使得機器能夠依據文章的內容，將文章分類為哪一種類別的新聞。如此一來，在校稿過程中如果發現有未標記或標記錯誤的問題，就可以快速標記正確的新聞分類，以減少人力及時間的耗費。此外，也可以為社群網站建立自有的新聞分類系統，將來自不同媒體的新聞資料，依據自有的分類方式進行分類，提供社群成員新聞資料。本論文分別利用爬蟲(web crawler)技術、資料前處理、結巴中文斷詞法來訓練電腦。經過多次的訓練及大量的訓練資料，實驗結果顯示新聞分類的準確率為97.42%。

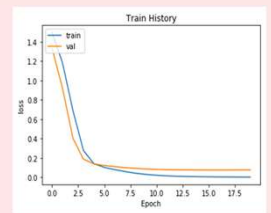
系統架構圖



NLP模型



模型Acc變化



模型Loss變化

網路介面



結論

AI在近幾年來迅速崛起，每個行業都想利用深度學習的技術，來提升各行業的營利，這也是讓深度學習快速崛起的原因。本專題分別利用爬蟲(web crawler)技術、資料的前處理、結巴中文斷詞法來訓練電腦。經過多次的訓練及大量的訓練資料，實驗結果顯示新聞分類的準確率為97.42%。除了可以協助新聞從業人員做新聞分類判斷，減少人力消耗之外；亦可以為社群網路提供自有的新聞分類系統，將來自不同媒體的新聞資料分類，更精準地推薦相關新聞給使用者。

