

摘要

隨著互聯網用戶數量的增加，在生活中無時無刻不在閱讀資訊和分享資訊。網路的發展給人們增加許多便利，但在螢幕下也有很多可信度低、惡意或不實的訊息稱為「假新聞(Fake News)」。2020年 COVID-19 病毒肆虐全球，相關的假新聞在全球範圍內傳播造成混亂、恐慌甚至死亡。因此，假新聞預測是非常必要的，本文希望設計一種有效的分類器，根據新聞摘要快速區分新聞是否為假新聞。本研究對 COVID-19 新聞數據集進行真假新聞預測，使用自然語言處理(Natural Language Processing, NLP)技術結合，深度學習 BERT、LSTM 和 BiLSTM 模型和機器學習 Logistic Regression、XGBoost、SVM 和 Random Forest 模型對資料集進行訓練，並比較模型績效。後續研將資料集分成 50%、25% 及 10% 的假新聞比例，研究結果表明 BERT 在這三種狀況下準確率為 0.97 都能達到很好的預測能力，證明 BERT 可以在資料量極度不平衡的狀態下就能進行學習。

關鍵字：COVID-19、假新聞、機器學習、BERT、LSTM