

摘要

科技日新月異，人們也越來越依賴電腦科技，但現有的搜尋引擎都需要對搜尋出的網頁進行仔細瀏覽，以去除不需要的資訊並保留需要的資訊，因此問答系統(**Question Answering System, QA System**)正在成為未來網路搜索的典範，它可以針對使用者所詢問的問題去找尋適合回覆的答案，以節省時間成本進而提高整體工作效率。因此本研究對問題分類進行探討，建置相關系統來準確回覆使用者，實驗中使用預訓練模型 **BERT**、**RoBERTa**、**RoBERTa_wwm** 及 **FastBERT** 對資料提取語意特徵，研究資料來自某電子公文公司，此研究對公司所提供之常見問答資料集進行前處理，先將問題中的標點符號、數字及停用字去除，接著使用來自 **Transformers** 雙向編碼向量 (**BERT**)、穩健優化的 **BERT** 預訓練方法 (**RoBERTa**)、經全詞遮罩的 **RoBERTa** (**RoBERTa_wwm**) 及具有自適應推理時間的自蒸餾 **BERT** (**FastBERT**) 對前處理後及原始問題進行特徵萃取及問題分類，接著對以上模型使用 **OPTUNA** 進行超參數優化，並與未進行優化的績效進行比較，以找最佳的參數組合。最後以 **Accuracy**、**Micro F1-score** 及 **Macro F1-score** 評估分類績效，結果顯示最佳配置為對原始問題建置 **RoBERTa** 以及使用前處理資料建立 **RoBERTa_wwm** 進行特徵提取分類 **Accuracy** 以及 **Micro F1-score** 為 **88.33%** 和 **88.45%**，而在 **Macro F1-score** 部分則是使用原始資料建置 **RoBERTa_wwm** 以及使用前處理資料建置 **BERT** 績效為 **79.06%** 和 **80.35%**，可以較準確地回覆使用者問題，此研究可給予使用傳統客服公司建置自動問答系統之參考，提供自動化客服機制，在最短時間內實現使用者需求且產生準確輸出。

關鍵字：問題分類、預訓練模型、超參數優化、**OPTUNA**