

經濟部 107 年度
人工智慧產業關鍵技術拔尖計畫 (1/4)
合作研究計畫

中文對話句法關聯分析之技術與研究
建議書徵求文件

財團法人資訊工業策進會

中華民國 107 年 02 月 26 日

107年度合作研究計畫建議書徵求文件

一、 簡介

高準確度的中文自然語言理解為建構一個優良的中文目標導向對話系統相當重要的基礎。由於中文文詞之間並沒有明確的分隔符號，因此進行中文自然語言理解之前，必須對於輸入文句進行中文句法關聯分析，其中包含中文斷詞、詞性標註及句法剖析等。錯誤的句法分析結果將使得對話內容與訓練資料集的句子難以對應，造成對話系統的自然語言理解準確率受限。目前離線版開源中文斷詞系統準確度過低，而較高準確度的CKIP必須付費商用且無法頻繁使用。因此將透過發展離線中文句法關聯分析技術，使繁體中文句法分析準確度高於離線版開源系統，達到提升自然語言理解準確度之效果。另外，支援離線使用及技術完全自主將可提升計畫成果可用性與效能，提高技轉成功的機會。

二、 計畫目標

發展準確度高於離線版開源系統之離線版繁體中文句法分析技術，透過對話中句法關聯分析的高準確度，回饋以提升主計畫於自然語言理解的準確度。此外，支援離線使用及技術完全自主提升計畫成果可用性與效能，提高技轉成功的機會。

三、 計畫範圍

1. 中文對話句法剖析功能雛形
 - 建構：中文對話句法剖析功能雛型
 - 驗證：研發中文對話句法剖析功能雛型，可以對於斷詞後的結果進行句法剖析，至少在一個公開測試集上驗證，如Chinse TreeBank(CTB)。

四、 預期成果

- 完成中文對話句法剖析功能雛型

※前述成果如有專利構想或專利申請產出時，需注意專利申請之新穎性(novelty)。因凡經公開發表之研發成果，如擬申請專利，須於公開發表後6個月內完成，前述成果如是以論文方式公開發表，將無法取得大陸與歐盟等國之專利。

五、 執行方式

- 合作計畫執行單位需定期舉行計畫控管會議，確保計畫的進度及執行方向一致性。
- 合作計畫執行單位需於6、12月繳交研究報告共二份。本單位要求合作計畫執行單位完成中文對話斷詞、詞性標註及句法剖析功能雛型。

六、 計畫期程及預估計畫總經費

計畫執行區間：107年01月01日至107年12月15日

總經費：600,000元

七、 驗收標準

- 完成中文對話句法剖析功能雛型

- 2018/06/30 交付期中報告1份，內容應包括：研究動機、架構規劃、演算法設計與分析、使用技術、文獻搜尋
- 2018/12/15 交付期末報告1份，內容應包括：研究動機、架構規劃、使用技術、實驗設計、模擬數據結果、結論、文獻搜尋，演算法驗證與測試

八、技術能力需求(請詳述所需要之技術能力或專長)

- 具自然語言處理研究經驗與成果之技術研究人員。

附件1：契約書格式

1-1：計畫書格式

1-2：經費動支報表

1-3：成果報告撰寫須知

1-4：報告格式

1-5：論文格式

1-6：保密聲明書

1-7：委託匯款同意書