

經濟部 109 年度
《人工智慧產業關鍵技術拔尖計畫(3/4)》
合作研究計畫

《知識圖譜時序動態關聯預測技術》
建議書徵求文件

財團法人資訊工業策進會

中華民國 109 年 02 月 06 日

109年度合作研究計畫建議書徵求文件

一、簡介

隨著企業自動化建立知識圖譜技術提升，整合其中所包含的全面性資料與社群網路中豐富的使用者互動記錄，並應用於各樣查詢檢索服務成為研發焦點。然而，由於知識圖譜本身具有動態更新的特性，其查詢檢索結果也將隨著動態知識圖譜或使用者的查詢偏好而發生變化。以事件規畫為例，利用知識圖譜與社群網路，將可由查詢獲得適合的參與者及感興趣的內容，但受限於當下所知的知識圖譜與社群網路，過去所發展的考慮參與者間社群連結的參與者推薦方法，只能以當下相關性來推薦參與者，無法推論未來知識圖譜並推薦內容；而現有的內容推薦方法則是根據當下相關性來預測內容喜好，亦無法推論整個知識圖譜並推薦參與者。

因此，我們需要發展對知識圖譜上的動態關聯預測技術，以推論未來知識圖譜，並據此推薦參與者及內容。現有的方法主要分為兩類：(1)轉譯距離模型 (Translational Distance Models)：將節點投射到特定關聯性的空間上學習嵌入向量 (embedding)，再計算嵌入向量之距離度以預測此關聯性存在之機率；(2)含義相配模型 (Semantic Matching Models)：以實體間與關聯的潛在特徵之相似度來判斷此關聯性存在之機率。然而，上述兩類關聯預測方法只依現況預測是否有關聯，無法針對所指定的時間來預測，更無法預測已存在的關聯是否可能消失 (或強度發生變化)。

另一方面，近年來深度學習逐漸受到重視，在圖結構上的研究可分為以下三類：(1)圖神經網路 (Graph Neural Network, GNN)：最初研究為循環圖神經網路 (RecGNN)，旨在假設節點經常與鄰近節點交換資訊 (直至達成穩定平衡為止)，運用循環神經網路的結構可習得節點嵌入向量；隨後亦有運用門控循環單元 (Gated Recurrent Unit, GRU) 延伸成循環圖神經網路的輸出序列的研究，可獲得更佳的效能。(2)圖卷積神經網路 (Graph Convolutional Network, GCN)：將原本運用於格狀結構資料上的卷積運算推廣至圖結構，不同於RecGNN，GCN疊加了多個圖卷積層以擷取高階節點特徵表示。(3)圖自編碼器 (Graph Autoencoders, GAEs)：主要用於學習節點嵌入及圖的生成分佈等，乃是非監督式學習的方法，將節點/圖編碼成一隱式向量空間，並以編碼資訊重建圖形資料。然而，上述的深度學習方法仍無法預測已存在的關聯是否可能消失或發生強度變化，因此，為了因應動態時序性知識圖譜與社群網路上的查詢應用，應提供支援動態關聯預測之技術，包含關聯的存在與否及強弱變化，以推論未來知識圖譜及回應查詢，即為本計畫的研發重點。

二、計畫目標

為辨識企業圖譜與外部圖譜是否可串接，基於同時時間點狀態下，節點之間的關聯性存在與否及強弱進行測量。但依據實際遇到的狀態，其實節點間是否存在關聯及其強弱，是會隨著動態知識圖譜或使用者的查詢偏好 (如使用者對地理位置的需求、或有特定的時間考量等) 而發生變化。

本計畫目標為未來事件規畫，其將整合社群網路、時序性知識圖譜及互動記錄，推論未來知識圖譜，並據此推薦參與者及事件內容。具體而言，本計劃所發展之技術將解決下列問題：

- (1)異質性網路節點表達：延伸Skip-Gram模型至異質性網路，學習具有meta-graph語意的嵌入向量。
- (2)維持一致性：隨著知識圖譜動態變化，嵌入向量除了需表達節點於當下圖譜結構之特性，也應維持該節點於不同時間點之一致性。

(3)預測關聯存在之強弱程度：以圖神經網路之技術輔以數學模型以預測其強弱程度，如時序點過程 (Temporal Point Process) 等。

本計劃之技術優勢在於比一般啟發式方法可考慮更多的隱性因素，無需仰賴領域專家所提供的規則；相較於傳統機器學習，本計劃技術無需人工擷取特徵，可維持圖譜結構，避免前處理時造成之資訊遺失，更可透過圖神經網路中多個隱藏層 (hidden layers) 考量特徵間的交互影響。

三、 計畫範圍

本計畫預計以企業的產品領域知識網路為資料來源，串接外部圖譜發展動態關聯預測技術，並應用於查詢檢索，工作項目包含：

- (1)整合時序性知識圖譜與社群網路為異質性網路，進行基於此網路的動態關聯預測技術與查詢系統的相關研究。
- (2)延伸Skip-Gram模型至時序性異質性網路，學習具有meta-graph語意且維持一致性的低維度嵌入向量。
- (3)以圖神經網路技術輔以數學模型 (如時序點過程) 以預測關聯之強弱程度。

四、 預期成果(明確說明合作研究成果之產出)

本研究計畫預期產出成果如下：

- (1) 知識圖譜時序動態關聯預測技術，將知識圖譜串接外部圖譜 (可結合社交、時間、空間等因素考量)，並支援動態關聯識別或測量，提升推論與查詢之成效。
- (2) 知識圖譜時序動態關聯預測技術實作程式，未來將可應用於電子商務平台之行銷，如PChome、momo購物網、生活市集、博客來、UDN shopping等，可推論未來知識圖譜和社群網路，並據此為相關消費者推薦適合商品，以協助規畫行銷策略。
- (3) 研究報告一份。

※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(包括計畫時程、計畫分工方式)

查核點	時間	產出物	報告架構/大綱
M1	2020/06/15	期中報告	<ul style="list-style-type: none">■ 知識圖譜時序動態關聯預測技術報告書■ 文獻探討■ 建立知識圖譜時序動態關聯語料■ 知識圖譜時序動態關聯預測演算法設計
M2	2020/11/30	研究報告 程式雛形	<ul style="list-style-type: none">■ 知識圖譜時序動態關聯預測技術期末研究報告書 (期末報告需包含雛形軟體與既有技術之實驗比較)■ 雛形軟體成果與操作說明

六、 計畫期程及預估計畫總經費

計畫執行區間：109年01月01日至109年12月31日

總經費：800,000元

七、 驗收標準(含教育訓練)

- (1) 知識圖譜時序動態關聯預測技術實作程式(包含雛形軟體與既有技術之實驗比較，包含嵌入向量學習及關聯存在/強弱預測模型，特定領域知識圖譜及目標文本相關語料)。
- (2) 研究報告一份。

八、 技術能力需求(請詳述所需要之技術能力或專長)

- 熟悉知識圖譜查詢相關技術。
- 具備圖查詢演算法設計經驗，熟悉Temporal Knowledge Graph (TKG)、Deep Temporal Reasoning技術，可進行動態關聯預測研究、設計對應檢索演算法，及研究分析能力。
- 具備程式撰寫與軟體架構開發能力，實作演算法原型，並進行實驗量化驗證。

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