

葉鎮源、楊維邦、柯皓仁、鄭培成 (2014), 『應用語句關係網路計算語句向心性之新聞事件摘要方法』, *資訊管理學報*, 第二十一卷, 第三期, 頁 271-304。

應用語句關係網路計算語句向心性之新聞事件摘要方法

葉鎮源*

國立自然科學博物館營運典藏與資訊組

楊維邦

國立東華大學資訊管理學系

柯皓仁

國立臺灣師範大學圖書資訊學研究所

鄭培成

健行科技大學資訊管理學系

摘要

摘錄式摘要技術的核心在於評估語句的摘要代表性，藉以排序語句作為摘錄語句時的依據。本研究將語句視為節點，藉由語句相似度來決定節點間是否存在連結，依此建構出語句關係網路模型。接著，衡量節點在網路中的重要性或對於其他相連節點的影響性，提出：(1) Degree Centrality (2) Normalized Similarity-based Degree Centrality、(3) HITS Centrality、(4) PageRank Centrality，及(5) iSpreadRank Centrality 的節點向心性分析；並以語句向心性作為語句的摘要代表性，藉此達到排序語句的目的。最後，導入 CSIS (Cross-Sentence Information Sub-sumption) 過濾重複性資訊，依序擷取語句組成摘要。實驗使用 DUC 2004 資料集來驗證上述摘要方法的可行性。在 ROUGE-1 的指標下，結合不同語句向心性之摘要效能依序是：iSpreadRank > Normalized Similarity-based Degree > PageRank > Degree > HITS。整體而言，實驗得知應用語句關係網路計算語句向心性之摘要方法確實可行。

關鍵詞：多文件摘要、摘錄式摘要、語句關係網路、網路節點向心性、語句排序

* 本文通訊作者。電子郵件信箱：jenyuan@mail.nmns.edu.tw
2013/09/29 投稿；2014/03/18 修訂；2014/04/26 接受

Yeh, J.Y., Yang, W.P., Ke, H.R. and Cheng, P.C. (2014), 'Extraction-based News Summarization Using Sentence Centrality in the Sentence Similarity Network', *Journal of Information Management*, Vol. 21, No. 3, pp. 271-304.

Extraction-based News Summarization Using Sentence Centrality in the Sentence Similarity Network

Jen-Yuan Yeh*

Department of Operation, Visitor Service, Collection and Information Management,
National Museum of Natural Science

Wei-Pang Yang

Department of Information Management, National Dong Hwa University

Hao-Ren Ke

Graduate Institute of Library and Information Studies, National Taiwan Normal
University

Pei-Cheng Cheng

Department of Information Management, Chien Hsin University of Science and
Technology

Abstract

Purpose: One widely-adopted summarization paradigm, sentence extraction, aims at extracting important sentences and composing them into a summary. The foundation towards sentence extraction is to assess importance of sentences in the summary so as to rank sentences for extraction. This paper employs graph-based text analysis to model documents and investigates measures of graph-based centrality as sentence salience in summarization.

Design/methodology/approach: This paper models documents on the same (or related) topic as a sentence similarity network, in which a sentence is regarded as a node and relationship between sentences only exists if they are semantically related. Several

* Corresponding author. Email: jenyuan@mail.nmns.edu.tw
2013/09/29 received; 2014/03/18 revised; 2014/04/26 accepted

methods for evaluating the importance of a node (i.e., a sentence) in the network are then proposed, namely: (1) Degree Centrality; (2) Normalized Similarity-based Degree Centrality; (3) HITS Centrality; (4) PageRank Centrality; and (5) iSpreadRank Centrality. All are designed on the basis of the idea that the importance of a node is determined not only by the number of nodes to which it connects, but also by the importance of its connected nodes. As to summary generation, CSIS (Cross-Sentence Information Sub-sumption) is employed for anti-redundancy while extracting sentences according to the sentence ranking produced based on the centrality of sentences.

Findings: The proposed summarization method was evaluated using the ROUGE evaluation suite on the DUC 2004 news stories collection. Experimental results show that, while considering the ROUGE-1 metric, the performance ranking is: iSpreadRank > Normalized Similarity-base Degree > PageRank > Degree > HITS. Another experiment, conducted to combine sentence centrality with surface-level features, also presents competitive results, compared with the best participant in the DUC 2004 evaluation.

Research limitations/implications: Directions for future research would be: (1) instead of symbolic-level analysis, to take into account semantics, such as synonymy, polysemy, and term dependency, while determining if two sentences are semantically related; (2) to investigate graph-based centrality developed in social network analysis for evaluating sentence salience in summarization; (3) to improve the cohesion and coherence of summaries using natural language processing techniques, such as sentence planning and generation.

Practical implications: The proposed summarization method is in an unsupervised manner; thus no training dataset is required. Since no domain-specific knowledge or deep linguistic analysis is exploited, the method is domain- and language-independent. However, it might lead to poor understanding of the input texts and would probably produces poor summaries, due to neither deep analysis of natural language processing performed, discourse structure considered, nor domain-specific knowledge involved in the process of summarization,

Originality/value: The contributions of this work are threefold. First, this paper offers a sentence similarity network to model topic-related documents. Second, novel

graph-based sentence ranking methods are explored to rank the importance of sentences for extraction. Finally, the proposed method had been proven successful in a case study with the DUC 2004 benchmark dataset.

Keywords: multidocument summarization, extraction-based summarization, sentence similarity network, network-based sentence centrality, sentence ranking

壹、前言

在資通訊科技蓬勃發展的現代，網際網路已是生活中不可或缺的要素，更扮演帶動人類文明往新的資訊紀元邁進的一大推手。拜科技之賜，資訊的生產、保存、傳播、利用與服務不再侷限於傳統實體媒體，數位化使得資料能被永久保存以供隨時隨地取用；透過網際網路，數位內容亦可無遠弗屆地傳遞分享到世界上各個角落。數位化技術的成熟，再加上近年來雲端計算（Cloud Computing）相關應用的實現，不但促使資料量快速倍增，同時也改變人們對於資料擷取的方式和處理速度。巨量資料（Big Data）使得搜尋及辨識資訊的困難度大幅提高，直接影響到資訊處理的有效運作與決策效率。如何快速、有效獲得真正符合需求的資訊，儼然已是目前資訊尋求急待解決的重要課題。資訊檢索（Information Retrieval）與文件摘要（Text Summarization）被認為是解決前述困境的兩項關鍵技術（Gong & Liu 2001）。資訊檢索依據檢索條件（Query）進行資料庫中資訊的比對及篩選，找出相關的資料回應使用者的需求；文件摘要則是將相關的內容語意和主題資訊進行分析、統整，轉化成簡要的形式呈現，協助人們在最短時間內掌握資訊的意涵，藉以判斷是否符合需要。

文件摘要，顧名思義，是由文件中擷取出重要資訊的過程，其產出是原始文件的精簡化版本，足以取代原文作為人們或應用系統的判斷和決策依據（Mani & Maybury 1999）。換言之，乃是從文件中萃取出重要的語意或主題資訊，再以簡明的摘要呈現，讓讀者能夠迅速理解原文中所要傳達的重點與主旨。文件摘要的類型大略可依文件來源、形成方式、摘要目的與讀者需求等四個面向來討論（葉鎮源 民 91）。首先，依輸入文件的數量分為單文件摘要（Single Document Summarization）與多文件摘要（Multi-document Summarization）。前者將單篇文件的內容精簡化、重點化，著重於有效擷取出語意及主題內涵；後者將多篇主題相關¹的文件融合後進行濃縮，另需考慮相關文件的資訊重複性。其次，摘要的形式有摘錄式摘要（Extract）和概述式摘要（Abstract）。第一種自原文中抽取出具代表性的關鍵詞、語句或段落；第二種則是萃取原文的語意或主題資訊，再經由資訊融合及語言生成進行重新編寫。再者，摘要內容通常有指示（Indicative）的目的，協助讀者區別原文是否符合需求；或是資訊性（Informative），提供與原文相同的主題資訊；或具有評論性質（Critical），呈現有別於原文的不同觀點論述。最後，一般性（Generic）摘要用於呈現原文所涵蓋的重要主題，而以使用者需求為導向（User-Oriented）的摘要則僅提供與查詢相關的內容。

以新聞事件來說，各家報社從不同的觀點和角度來陳述相同的事件；這些主

1 多篇文件內容所涵蓋的主題相似或重複時，則稱這些文件為主題相關（Topic-related）。

題相關的報導動輒數十篇，甚或是近百篇，唯有投入時間多方閱讀比較，方可掌握完整的新聞事件發展脈絡。有幸的是，藉由多文件摘要技術的應用，將報導相同事件的新聞編製成摘要，即可幫助讀者快速得知事件發展的來龍去脈。本文提出應用語句關係網路計算語句向心性之新聞事件摘要方法：首先將語句視為節點，透過語句的相似度決定語句間是否存在有語意連結，用以建構語句關係網路作為新聞事件的主題模型。接著，衡量節點在網路中的重要性或對於其他相連節點的影響性，提出：(1) Degree Centrality、(2) Normalized Similarity-based Degree Centrality、(3) HITS Centrality、(4) PageRank Centrality，以及(5) iSpreadRank Centrality 共五種節點向心性的分析，並將語句向心性當作是決定語句摘要代表性的依據，藉此排序語句作為摘要生成時語句擷取順序的參考。最後，導入 CSIS (Cross-Sentence Information Sub-sumption) (Radev et al. 2004) 過濾重複資訊；在摘要長度的限制下擷取出適量的語句，再依語意的連貫性重組後輸出摘要文本。

本文共六個章節：第貳節介紹文件摘要相關的研究；第參節詳述應用語句關係網路計算語句向心性之新聞事件摘要架構；第肆節說明實驗設計及結果討論；第伍節探討所提出之摘要方法的特性與貢獻；最後是結論和未來研究發展建議。

貳、相關研究

文件摘要的起源最早可追溯到 1950 年代 (Luhn 1958)；然而其至今在技術上的發展，絕大多數仍遵循著傳統的典範，那就是計算語句的資訊量多寡作為語句的重要程度，再依照語句的重要性摘錄出足以涵蓋原文主旨的語句或段落。文件摘要的發展，大致上可區分為五個階段 (Yeh et al. 2005)：

1. 1950 至 1960 年代，研究方法著重於寫作型態 (Genre) 的分析，主要是運用詞頻、提示字詞和語句位置等特徵衡量語句所提供的資訊量多寡。舉例來說，段落的第一句通常點明接下來所要描述的主題，或是語句中出現特定的提示片語 (例如：“in summary”或“in conclusion”)。這類方法的優點是實作容易，但因其與文件類型相關，導致技術的再利用性並不高；
2. 1970 至 1980 年代，研究方法轉為以建構知識模型來解析文件的意涵。一般是以框架 (Frame) 或模板 (Template) 的輔助來擷取人物、地點及時間等要素，再經由知識模型的推導產出摘要。缺點是知識模板需由專家人工定義，而模板定義的詳盡度不足時，可能導致摘要呈現的事實或語意與原內容有落差；
3. 1990 年代開始，資訊檢索被廣泛地應用；相較於文件檢索來說，文件摘要是檢索出與主題相關的語句。然其分析多半著重在字詞的層面，未考慮同義詞、一詞多義或字詞依屬關係之語意層級的解析，導致無法產出適當的

摘要內容；

4. 1960 至 1970 年代，以及 1970 至 1980 年代左右，分別同時有受到有計算語言學及認知心理學影響而發展出來的摘要技術；
5. 自 2000 年開始，由於搜尋引擎的快速發展，網頁結構的圖形模型分析開始受到矚目，間接影響到文件摘要之主題模型的表達及語句排序等研究。

本研究屬於多文件摘要的範疇。一般認為多文件摘要的研究先驅非 McKeown 與 Radev (1995) 莫屬；他們使用模板成功擷取出人物、地點及時間要素，再經過語句生成產生摘要。類似的研究還有：Harabagiu 與 Maiorano (2002)、Ji 等 (2013)。事實上絕大多數的多文件摘要方法利用分群 (Clustering) 的技術將語句或是段落進行分群，所得到的群集定義為文件集中所要傳達的內容主題 (Theme)；相關的研究有：Aliguliyev (2010)；Barzilay、McKeown 與 Elhadad (1999)；Daniel、Radev 與 Allison (2003)；Goldstein 等 (2000)；McKeown 等 (1999)；Sathish kumar 與 Sharmila (2013)；Xia、Zhang 與 Yao (2011)。考量群的大小，當群擁有的語句越多時，代表此群越能代表整個文件集。因此，摘要的產生可依據群的大小依序從各個群中挑選出最具代表性的語句。值得一提的是，因為分群的排他性，由各個群中所挑選的語句，其彼此的資訊重複性低，此點符合多文件摘要須過濾重複資訊的特性。此外，傳統單文件摘要中使用語句特徵來衡量語句重要性的方法亦被廣泛地應用在多文件摘要；例如：Gupta、Chauhan 與 Garg (2012)；Li 與 Li (2013)；Radev 等 (2004)；McDonald 與 Chen (2006)；Suanmali、Salim 與 Binwahlan (2011)。

一、基於圖形結構之文件摘要技術

Salton 等 (1997) 以段落為節點，計算兩兩段落的相似度而建構出主題關係地圖 (Text Relationship Map)，可稱得上是以圖形結構為基礎之摘要方法的始祖。他們假設當節點所擁有的連結數目越多時，表示該段落與整篇文件的內容主題越相近；根據連結數目的多寡決定段落的重要順序，藉此摘錄段落輸出摘要。Mani 與 Bloedorn (1999) 的研究將關鍵字詞視為節點，藉由解析語句中隱含的字詞關係，包括有：(1) 片語關係 (PHRASE)、(2) 形容詞關係 (ADJ)、(3) 同義關係 (SAME)，及 (4) 關聯關係 (COREF) 等，進而將文件轉化成關鍵字詞語意關聯的圖形模型 (如圖 1 所示) 接著，比較兩兩文件模型的相似性 (Commonality) 和差異性 (Difference) 得到「相似」、「差異」兩組關鍵字詞集合；最後，利用關鍵字詞詞頻評估語句的代表性，再從相似與差異兩個面向挑選出組織摘要的語句。

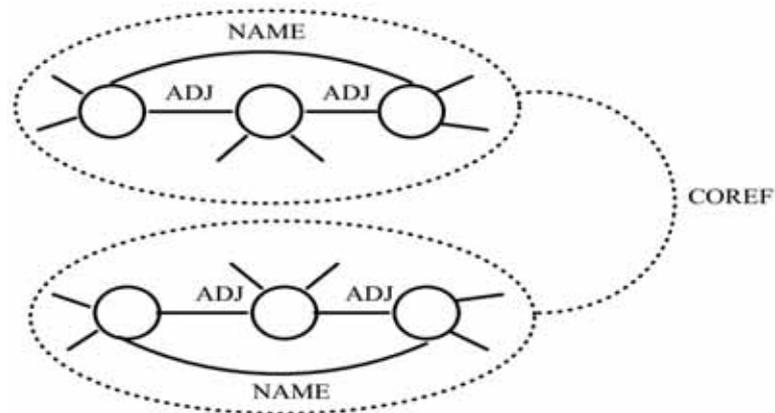
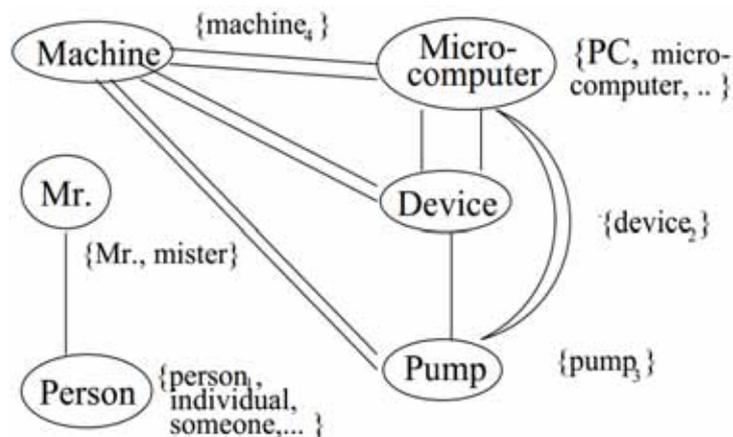


圖 1：關鍵字詞語意關聯模型示意 (Mani & Bloedorn 1999)

另一個代表是：Barzilay 與 Elhadad (1997) 使用語彙鏈結 (Lexical Chains) 萃取文件中的概念。語彙鏈結是文件中具有相同意義或關聯的字詞所構成的集合，如圖 2 所示；利用 WordNet²中定義的字義與字詞關係決定語彙連結的類別，包括：(1) Extra-Strong (定義字詞與其同義字詞的關係)，(2) Strong (定義兩個字詞在 WordNet 中存在直接關聯的關係)，及(3) Medium-Strong (定義兩個字詞在 WordNet 中存在間接關聯的關係)。根據長度和同質性 (Homogeneity) 評比語彙鏈結的重要性，藉此挑選出代表性的語彙鏈結；最後，依照語彙鏈結中的概念字詞挑選有出現這些字詞的語句形成摘要。



說明：{Mr., Person}表示「人」的概念，{Machine, Micro-computer, Device, Pump}表示「機器」的概念。

圖 2：語彙鏈結範例 (Barzilay & Elhadad 1997)

2 <http://wordnet.princeton.edu/>.

過去以圖形結構為模型的研究未能引起共鳴，其原因歸咎於圖形模型的建構過程繁瑣不易；若採用字詞關係的解析，需要藉由領域知識 (Domain Knowledge) 或字典 (Thesaurus) 的輔助，但這類資源常因領域的差異而難以取得。近年來由於 PageRank (Brin & Page 1998) 與 HITS (Kleinberg 1999) 的問世，運用網路結構來排序網頁的重要性；在文件摘要的領域中，已有學者引用 PageRank 與 HITS 來增進文件摘要的效能和正確性。例如，LexRank (Erkan & Radev 2004) 以 PageRank 為多文件摘要的核心模型，進而賦予語句重要性以標示語句為摘要與非摘要語句。Erkan (2006) 擴充 LexRank，利用個人化排序的 Biased PageRank (Haveliwala 2002) 研發問答式多文件摘要系統。Mihalcea (2004) 及 Mihalcea 與 Tarau (2005) 實驗不同的圖形排序演算法，並驗證這類演算法在單文件及多文件摘要的可行性。Wan 與 Yang (2008) 認為語句群集背後所代表的主題重要性不應視為相同，以群集為對象，分析群集的關係並結合 HITS 提出多文件摘要方法。Zhang、Sun 與 Zhou (2005) 使用 HITS 作為語句排序的核心，同時整合語句特徵提出更優良的多文件摘要架構。其他相關研究請參考：Boudin、Huet 與 Torres-Moreno (2011)；Cai 與 Li (2013)；Canhasi 與 Kononenko (2011)；Christensen 等 (2013)；Wan 與 Yang (2006; 2008)；Zhu 等 (2007)。

二、多文件摘要技術的比較

常見的多文件摘要技術類型，包括有：知識模板萃取、主題偵測分群、語句特徵分析，以及近年興起的基於圖形結構的摘要方法 (Das & Martins 2007; Nenkova & McKeown 2012)。知識模板萃取擷取人物、地點及時間等要素，再經由知識模型的推導產出摘要；其優點是結合領域知識進行較準確的文件分析，但缺點是知識模板需由專家人工定義，而模板定義的詳盡度不足時，可能導致摘要呈現的事實或語意與原內容有落差。

主題偵測分群的優點是藉由段落或語句的分群，達到將隱含不同主題的段落或語句進行分類的目的；而群集間的不相關性，只需從每個群集擷取代表的段落或語句，通常不必再處理重複資訊的過濾。然而，其缺點是分群結果好壞直接影響到摘要品質；另外，一般需要事先定義產出群集的數量，而實務上群集數目通常難以在最初時便可決定。

語句特徵分析的好處是特徵擷取方式簡單、容易實作，但因其著重個別語句的分析而缺乏與其他語句的關聯性考量。這一點在多文件摘要的應用上易顯不足：語句特徵的計算侷限於單篇文件的範圍，例如，語句位置僅考慮單篇文件中的位置，可能發生來自不同文件的語句，其特定特徵值在不同基礎下無法有效鑑別的情形。

基於圖形結構的摘要方法將文件結構轉換為圖形網路模型，其優點是考慮單一文件內的語句或段落關聯之外，同時納入跨文件(Cross-Document)的關係建構，因而提供語句或段落關係之整體性分析的可能；並可藉助圖形結構理論中用來描述與量化網路結構的方法衡量節點在網路結構中的重要性或對於其他相連節點的影響力。其次，部分圖形結構分析技術具有半監督式學習(Semi-supervised Learning)的效果，以得到較佳的節點向心性計算；例如：PageRank (Brin & Page 1998) 與 HITS (Kleinberg 1999)。最後，若將語句節點向心性視為特定的語句特徵，便可直接整合於語句特徵分析的摘要方法中；對於多文件摘要來說，圖形結構化的模型直接提供了在跨文件關係下進行特徵比較的基礎。

過去的文獻提出許多基於圖形結構的摘要方法，但多半只侷限於單一技術的應用，鮮少針對不同種類的節點分析之於段落或語句排序的影響及摘要成效進行綜合討論。有感於此，本文基於語句關係網路模型，導入圖形結構理論用來評估節點向心性的法則，提出五種語句向心性計算方法，同時利用語句向心性來決定語句的摘要代表性，藉以達到排序語句的目的；研究的目的除了應用語句向心性研發新型的語句排序與摘要技術外，同時希冀在相同的摘要架構下進行不同語句關係網路模型和不同語句向心性分析的各種組合應用於摘要生成的可行性探討及效能評估比較，藉以補足相關研究版圖的缺塊。這點也是本文有別於其他圖形結構相關研究的最大不同與貢獻。

最後，實務上認為文件摘要的技術各有其優缺點；考量應用領域的不同，以及摘要目的與輸出摘要類型的差異，很難用單一摘要技術得到不錯的成效。因此，近年來文件摘要技術的發展多以混合式(Hybrid)方法來進行；例如：Cai 與 Li (2013)；Wan 與 Yang (2008)。

參、應用語句關係網路計算語句向心性之新聞事件摘要

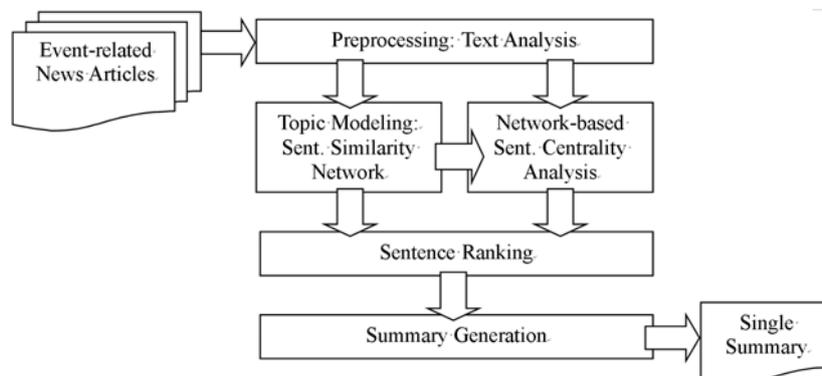


圖 3：應用語句關係網路計算語句向心性之新聞事件摘要架構

圖 3 是本文所提出的「應用語句關係網路計算語句向心性之新聞事件摘要」架構，其輸入是多篇陳述相同事件發展的新聞報導³，輸出為單篇摘要。整體架構共五個模組：

- 文本分析 (Text Analysis)：進行輸入文件的前置處理，包括：斷詞切字、語句辨別、詞性標記及詞頻計量；同時轉化語句為以關鍵字詞組成之語彙空間中的向量表示；
- 主題模型建構 (Topic Modeling)：以語句為節點，計算兩兩語句的關聯性；接著依據關聯強度決定任兩語句間是否存在有語意連結，依此建構出語句關係網路 (Sentence Similarity Network) 作為新聞事件的主題模型⁴；
- 語句向心性分析 (Sentence Centrality Analysis)：利用圖形結構理論中用來描述與量化網路結構的法則，衡量節點在網路的重要性或對於其他相連節點的影響，藉此計算節點之向心性來決定語句的重要程度；
- 語句排序 (Sentence Ranking)：根據語句向心性分析得到的語句代表性來排序語句；排序時由權重值高的排到權重值低的，用以作為摘要生成模組中語句擷取的依據；
- 摘要生成 (Summary Generation)：此模組包括語句擷取 (Sentence Extraction) 和摘要組成兩個部份。依據摘要的長度限制擷取出適當數量的語句，同時考慮語意的連貫性及內容組織的一致性，重組產生最後輸出的摘要。

一、文本分析

本研究依據資訊檢索的原則，從文件集中擷取出具有代表性的關鍵字詞當作建構語彙向量空間的維度特徵，進而將文件中的語句⁵轉化成為座落在相同語彙空間中的向量，使得任意兩個語句得以在共通的描述基底上進行差異或相似的量化比較，用以作為新聞事件主題建模中衡量語句關聯性的基礎。

文本分析的處理步驟，依次是：斷詞切字 (Tokenization)、語句邊界偵測 (Sentence Boundary Detection)、詞性標記 (Part-of-Speech Tagging)、名詞詞組化 (Noun-Phrase Chunking)、停用詞刪除 (Stopword Removal) 及詞幹還原 (Stemming)。實作時使用 LT-TTT2⁶進行各項處理，且參考 Snowball⁷停用詞表過濾停用詞，同時採用 Porter Stemming (Porter 1980) 進行詞幹還原。上述程序將輸

3 本文以新聞事件摘要為例，但所提出的摘要方法亦適用於其他主題相關之多文件的輸入。

4 本研究延伸 Salton 等 (1997) 的主題關係地圖，進而提出語句關係網路作為新聞事件的主題模型。

5 本研究以語句 (Sentence) 作為摘要系統中最根本的資訊處理單元。實作時亦可依實際的摘要需求，調整成關鍵字詞 (Word) 或片語 (Phrase)，或是段落 (Paragraph)，甚至文章片段 (Segment)。

6 <http://www.ltg.ed.ac.uk/software/lt-ttt2/>.

7 <http://snowball.tartarus.org/>.

入文件切分為語句和關鍵字詞兩個集合。考量絕大多數的語句是由「主詞(Subject) 述詞(Predicate) 受詞(Object)」的結構所組成(陳光華 1998)，其中主詞與受詞通常為名詞，述詞一般是動詞；因此，在實作時只保留名詞與動詞作為向量空間的語彙特徵。

接著，基於語句中出現關鍵字詞的詞頻，計算語句在相對空間維度的特徵權重，進而轉換語句成為語彙空間中的一組向量。給定語句集合是 $S = \{s_1, \dots, s_n\}$ ，關鍵字詞集合是 $W = \{t_1, \dots, t_m\}$ ，語句 $s_j = \{t_{j,1}, t_{j,2}, \dots, t_{j,p}\}$ 表示出現在 s_j 中的關鍵字詞集合 $W_j = \{t_{j,k} | k = 1, \dots, p\}$ 且 $W_j \subseteq W$ ；此時將 s_j 表示為如 Eq. (1) 所定義的語句向量：

$$s_j = \langle w_{1,j}, w_{2,j}, \dots, w_{m,j} \rangle \quad (1)$$

其中， $w_{i,j}$ 是 s_j 在空間維度 t_i 的特徵量化權重，可依據 TF-IDF(Salton & Buckley 1988) 計算得到，詳見 Eq. (2)：

$$w_{i,j} = \frac{tf_{i,j}}{\max_l tf_{l,j}} \times \log \frac{N}{n_i} \quad (2)$$

此處， $tf_{i,j}$ 是 t_i 出現在 s_j 中的次數， $\max_l tf_{l,j}$ 代表 s_j 中出現最多次數之關鍵詞 t_l 的次數。 $\log(N/n_i)$ 則是 t_i 的 IDF 因子，而 n_i 是擁有 t_i 的語句個數， N 是所有語句的數量。IDF 是反向文件頻率；當關鍵字詞 t_i 的 IDF 值越低時，說明 t_i 出現於越多的語句中，其重要性（或稱類別區分能力）相對而言就較差。

二、新聞事件主題模型：語句關係網路

本研究中將語句被視為網路模型中的節點 (Node)，衡量任意兩個節點的關聯強度可決定節點間是否存在有連結 (Link)，據此形成語句關係網路作為新聞事件的主題模型。通常，節點間的關聯強度定義為語句間的相似度；如 Eq. (3) 所示，計算語句 s_i 和 s_j 的向量夾角差異（即，餘弦相似度 Cosine Similarity）可作為此兩個語句的相似度。

$$sim(s_i, s_j) = \frac{\vec{s}_i \cdot \vec{s}_j}{|\vec{s}_i| \times |\vec{s}_j|} \quad (3)$$

以 s_i 和 s_j 來說，若相似度 $sim(s_i, s_j)$ 大於或等於事先定義的臨界值 α 時，則定義 s_i

和 s_j 間存在語意連結關係 (Semantic Related Link); 換句話說, 具有語意連結關係的兩個節點, 其相似度必定高於臨界值 α 。

假設 $V = S = \{s_1, \dots, s_n\}$ 是語句節點集合, $E = \{e_{ij} \mid e_{ij} \text{ 是 } s_i \text{ 與 } s_j \text{ 間的連結}; e_{ij} \text{ 存在, 若且唯若 } sim(s_i, s_j) \geq \alpha\}$ 是語句關係連結集合, 兩種不同類型的語句關係網路, 稱之為「連續權重語句關係網路 (G')」與「離散權重語句關係網路 (G'')」, 其定義詳如 Eq. (4) 及 Eq. (5) :

$$G' = (V, E')$$

- $E' = \{e_{ij} \mid e_{ij} \text{ 存在, 若且唯若 } sim(s_i, s_j) \geq \alpha\}$ (4)
- $\forall e_{ij} \in E'$, 定義 e_{ij} 的權重值 $w(e_{ij}) = sim(s_i, s_j)$; 反之則為 0。

$$G'' = (V, E'')$$

- $E'' = \{e_{ij} \mid e_{ij} \text{ 存在, 若且唯若 } sim(s_i, s_j) \geq \alpha\}$ (5)
- $\forall e_{ij} \in E''$, 定義 e_{ij} 的權重值 $w(e_{ij}) = 1.0$; 反之則為 0。

實作時將 α 設定成 0.1⁸。同時, 利用 $n \times n$ 的相鄰矩陣來表示 G' 與 G'' ; 由 $w(e_{ij}) = w(e_{ji})$ 可知此相鄰矩陣亦是對稱矩陣。值得一提的是, G' 與 G'' 都是無向權重圖 (Undirected Weighted Graph), 兩者最大的差別只在於 e_{ij} 的權重值不同。

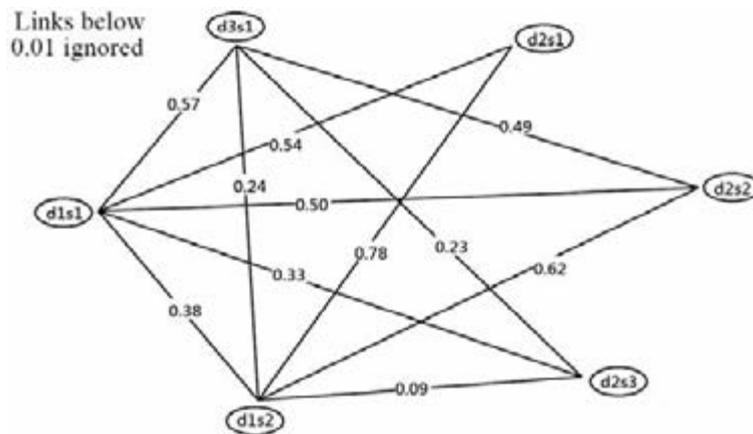


圖 4：語句關係網路模型範例

圖 4 是語句關係網路模型的範例; 節點 $d_i s_j$ 代表文件 d_i 的第 j 個語句, 臨界值 α 是 0.01。以 d_{1s1} 與 d_{3s1} 的來說, 兩者的相似度是 0.57, 大於臨界值 0.01, 因而存

8 α 值的大小決定語句關係網路的疏密程度: 當 α 值高時, 節點間的連結數目較少而形成稀疏網路; 相反地, 當 α 值低時, 節點間的連結數目較多而形成密集網路。通常來說, 網路的稀疏與密集直接影響到文件模型的資訊失真程度; 因此, α 值的設定需視實際應用經由實驗來決定。

在有連結關係；相反地， d_2s_2 和 d_2s_3 的相似度低於臨界值，所以在語句關係網路中不存在有連結關係。圖 5 為圖 4 的語句關係網路所相對應之 G' 與 G'' 的相鄰矩陣表示。

$$A' = \begin{matrix} & \begin{matrix} d_1s_1 & d_1s_2 & d_2s_1 & d_2s_2 & d_2s_3 & d_3s_1 \end{matrix} \\ \begin{matrix} d_1s_1 \\ d_1s_2 \\ d_2s_1 \\ d_2s_2 \\ d_2s_3 \\ d_3s_1 \end{matrix} & \begin{bmatrix} - & 0.38 & 0.54 & 0.50 & 0.33 & 0.57 \\ 0.38 & - & 0.78 & 0.62 & 0.09 & 0.24 \\ 0.54 & 0.78 & - & 0.00 & 0.00 & 0.0 \\ 0.50 & 0.62 & 0.00 & - & 0.00 & 0.49 \\ 0.33 & 0.09 & 0.00 & 0.00 & - & 0.23 \\ 0.57 & 0.24 & 0.00 & 0.49 & 0.23 & - \end{bmatrix} \end{matrix}$$

圖 5(a) : G' 的相鄰矩陣表示 A'

$$A'' = \begin{matrix} & \begin{matrix} d_1s_1 & d_1s_2 & d_2s_1 & d_2s_2 & d_2s_3 & d_3s_1 \end{matrix} \\ \begin{matrix} d_1s_1 \\ d_1s_2 \\ d_2s_1 \\ d_2s_2 \\ d_2s_3 \\ d_3s_1 \end{matrix} & \begin{bmatrix} - & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & - & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & - & 0.00 & 0.00 & 0.00 \\ 1.00 & 1.00 & 0.00 & - & 0.00 & 1.00 \\ 1.00 & 1.00 & 0.00 & 0.00 & - & 1.00 \\ 1.00 & 1.00 & 0.00 & 1.00 & 1.00 & - \end{bmatrix} \end{matrix}$$

圖 5(b) : G'' 的相鄰矩陣表示 A''

三、語句向心性分析

語句關係網路模型的建立，提供依據圖形結構理論來評估語句節點向心性的基礎。常見描述與量化網路結構的分析方法，包括：網路的大小 (Size) 和密度 (Density) 節點的強度 (Intensity) 相互性 (Reciprocity) 及向心性 (Centrality) 等 (謝仰哲 民 97)。而節點的向心性一般用來衡量節點在網路結構中的重要性或對於其他相連節點的影響性 (Carrington et al. 2005)；例如：連結向心性 (Degree Centrality)、近距向心性 (Closeness Centrality) 和參與向心性 (Betweenness Centrality)。

本研究以節點的向心性來決定語句的重要程度，用以作為語句的摘要代表性；同時設計五種節點向心性的計算方法，分別為：(1) Degree、(2) Normalized Similarity-based Degree、(3) HITS、(4) PageRank，及(5) iSpreadRank。這些方法的概念都是將節點的連結關係當成是評估節點重要程度的投票機制；舉例來說，若節點 A 存在有一連結到節點 B 時，則稱節點 A 投票給節點 B ，而節點的向心性乃由各個節點得到票數來決定。

為了方便說明起見，以下各小節中假設語句關係網路是 $G=(V, E)$ ； G 依實際情況請替換成連續權重語句關係網路 G' 或離散權重語句關係網路 G'' 。

(一) Degree Centrality

Degree Centrality 定義成節點與其他節點間的連結個數。當語句節點的 Degree Centrality 越高時，表示此語句和其他語句間存在的連結數目越多；換句話說，此語句與越多語句的相似度都高。此時，稱該語句和其他相關語句所表達的語意主

題相近，因而可將其視為是其他相似語句的替代摘要。考量到語句關係網路的連結具有權重，本研究將節點的 Degree Centrality 廣義定義為節點的所有連結之權重值總和。以語句 s_i 來說，根據 Eq. (6) 計算可得到 s_i 的 Degree Centrality：

$$\text{deg}(s_i) = \sum_{j:e_{i,j} \in E} w(e_{i,j}) \quad (6)$$

由 Eq. (6) 可知，考慮離散語句關係網路 G' 時， s_i 的 Degree Centrality 是 s_i 擁有的連結數目；考慮連續語句關係網路 G 時， s_i 的 Degree Centrality 等同 s_i 擁有連結的權重值加總。

(二) Normalized Similarity-based Degree Centrality

相對於 Degree Centrality 計算節點的連結權重總和，Normalized Similarity-based Degree Centrality 不但考慮單一連結的關聯強度，更進一步考量所有相連節點的關聯強度總和作正規化處理，定義如下：

$$\text{sim deg}(s_i) = \sum_{j:e_{i,j} \in E} \frac{w(e_{i,j})}{\sum_{k:e_{j,k} \in E} w(e_{j,k})} \quad (7)$$

由此可知， s_i 由 s_j 得到的投票權重與 s_j 所有連結的權重總和成反比；也就是說， s_j 的投票是依比例平均分配給所有連結指向的節點，而非只獨厚單一節點 s_i 。

以圖 6 為例， s_1 與 s_2 的 Degree Centrality 都是 0.6。考慮 Normalized Similarity-based Degree Centrality 時， $\text{sim deg}(s_1) = 0.67$ ， $\text{sim deg}(s_2) = 1.33$ ，說明 Normalized Similarity-based Degree Centrality 比 Degree Centrality 具有更好的節點向心性鑑別度。

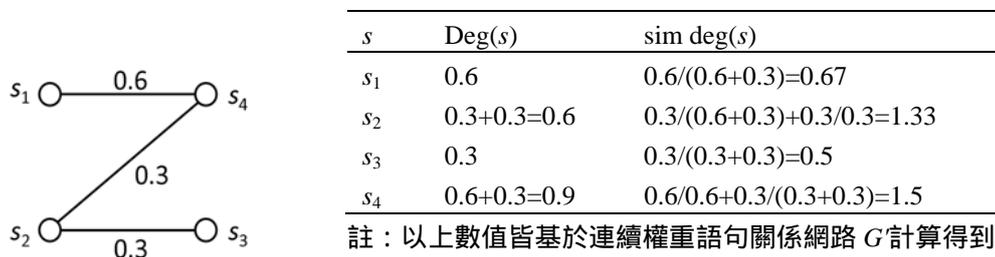


圖 6：語句關係網路的範例及其 $\text{deg}(s)$ 與 $\text{sim deg}(s)$ 的計算比較

(三) HITS Centrality

HITS (Hyperlink-Induced Topic Search) 是結合網頁正向連結 (Forward Link) 和反向連結 (Backlink) 結構分析進而決定網頁名次排行的方法 (Kleinberg 1999)。HITS 定義 Hub 及 Authority 兩種網頁類型作為網頁重要性排名的決策指標；Hub 是包含超連結至與主題相關的 Authority 網頁，而 Authority 是提供重要且可靠之主題資訊的網頁。HITS 假設一個好的 Hub 具有多個正向連結連到好的 Authorities，而一個好的 Authority 擁有許多好的 Hubs 連結到它；依此可以定義 Eq. (8) 和 Eq. (9) 計算網頁 p_i 的 Authority 與 Hub 指標權重：

$$\text{Authority Update Rule: } auth^{(k)}(p_i) = \sum_{p_j \text{ 有正向連結到 } p_i} hub^{(k-1)}(p_j) \quad (8)$$

$$\text{Hub Update Rule: } hub^{(k)}(p_i) = \sum_{p_i \text{ 有正向連結到 } p_j} auth^{(k)}(p_j) \quad (9)$$

實務上 HITS 建構節點的關係矩陣，並通過迭代 (Iterative) 計算每個節點的權威分數權重 (Authority Degree) 和匯集分數權重 (Hub Degree)。

本研究使用 HITS 來衡量節點的向心性。考量 HITS 所分析的網頁超連結結構具有方向性，因而假設語句關係網路中的連結為雙向連結以符合 HITS 的要求。以語句 s_i 來說，其 HITS Centrality 定義為 Eq. (10) 及 Eq. (11)：

$$\text{Authority Update Rule: } auth^{(k)}(s_i) = \sum_{j: e_{j,i} \in E} hub^{(k-1)}(s_j) \times \frac{w(e_{j,i})}{\sum_{k: e_{j,k} \in E} w(e_{j,k})} \quad (10)$$

$$\text{Hub Update Rule: } hub^{(k)}(s_i) = \sum_{j: e_{i,j} \in E} auth^{(k)}(s_j) \times \frac{w(e_{i,j})}{\sum_{k: e_{i,k} \in E} w(e_{i,k})} \quad (11)$$

其中， $auth^{(k)}(s_i)$ 與 $hub^{(k)}(s_i)$ 分別為 s_i 的 Authority 與 Hub 指標權重。作法上將節點的 Hub 指標權重初始設為 1.0，藉由不斷迭代運算的過程可得到所有語句最終的 Authority 與 Hub 指標權重。實作時採用語句的 Authority 指標權重當作語句的向心性。

最後要說明的是：考慮離散語句關係網路 G' 時，Eq. (10) 中 $\frac{w(e_{j,i})}{\sum_{k: e_{j,k} \in E} w(e_{j,k})}$ 表示

$hub^{(k-1)}(s_j)$ 平均分配給所有與 s_j 相連的語句節點; Eq. (11) 中 $\frac{w(e_{i,j})}{\sum_{k:e_{i,k} \in E} w(e_{i,k})}$ 表示 $auth^{(k)}(s_j)$

的分配與和節點 s_j 之連結個數成反比。而考慮連續語句關係網路 G 時, $hub^{(k-1)}(s_j)$ 和 $auth^{(k)}(s_j)$ 的分配是依照連結權重及相連節點之平均關聯強度比例來決定。

(四) PageRank Centrality

PageRank 是 Google 搜尋引擎用來決定網頁重要性, 據此將最相關且可靠度高的網頁呈現在搜尋結果頂端的技術 (Brin & Page 1998), 其概念是將網頁的超連結當作評估網頁重要程度的投票機制。給定網頁 T_1, T_2, \dots, T_n 是具有正向連結至網頁 A 的網頁集合, 此時 A 的 PageRank 值 $PR(A)$ 定義如下:

$$PR(A) = \frac{(1-d)}{N} + d \left(\frac{PR(T_1)}{C(T_1)} + \dots + \frac{PR(T_n)}{C(T_n)} \right) \quad (12)$$

其中, N 是所有網頁的個數; $PR(T_i)$ 是 T_i 的 PageRank 值; $C(T_i)$ 是 T_i 的正向連結數目。 $PR(T_i)/C(T_i)$ 代表 A 從 T_i 得到的 PageRank 值與其正向連結數目成反比; 即, T_i 的 PageRank 值是平均分配給所有正向連結所指向的網頁, 而非只獨厚網頁 A 。實數 d 稱之為阻尼係數 (Damping Factor), 其值介於 0 和 1 之間, 通常設定為 0.85。

PageRank 屬於隨機漫步 (Random Walk) 的使用者瀏覽行為模型。想像有個使用者正在隨機瀏覽某個網頁, 且經由網頁的正向連結不斷擴展開啟其他網頁, 而在某個時間點隨機跳到新的網頁, 之後繼續重複前述相同的動作; 此時, 網頁的 PageRank 值即是假設使用者瀏覽到某個網頁的機率。Eq. (12) 中的 $(1-d)/N$ 是該隨機瀏覽者在觀看某個網頁後, 決定不依循其正向連結連至其他網頁, 而是隨機開啟任意網頁進行瀏覽的可能性。

本研究引用 PageRank 計算語句節點的向心性。考量 PageRank 所分析的網頁超連結結構具有方向性, 此處假設語句關係網路中的連結是雙向連結, 使其符合 PageRank 的要求。有別於 PageRank, 本研究廣義地定義 Eq. (12) 中的 $C(T_i)$ 是 T_i 所有連結的權重值總和, 再考量單一連結的關聯強度, 最後定義語句 s_i 的 PageRank Centrality 為 Eq. (13):

$$PR(s_i) = \frac{(1-d)}{N} + d \sum_{j:e_{j,i} \in E} PR(s_j) \times \frac{w(e_{j,i})}{\sum_{k:e_{j,k} \in E} w(e_{j,k})} \quad (13)$$

其中, N 是所有語句的數量。實作中將每個節點的 PR 初始值設成 1.0, 藉由不斷迭代運算的過程可得到所有語句的最終 PR 值。

特別一提的是，考慮離散語句關係網路 G' 時，Eq. (13) 中的 $\frac{w(e_{j,i})}{\sum_{k:e_{j,k} \in E} w(e_{j,k})}$ 說明

$PR(s_j)$ 是平均分配給所有與 s_j 相連的語句節點；而考慮連續語句關係網路 G' 時， $PR(s_j)$ 的分配則是由連結權重及相連節點之平均關聯強度比例來共同決定。

(五) iSpreadRank Centrality

iSpreadRank 引用認知心理學中的擴散激發理論 (Collins & Loftus 1975; Quillian 1968)，假設與某個節點有連結的節點都會接受到由該節點所傳遞而來的不同程度影響；同樣地，這些受到影響而被激發的節點會接著傳遞並激發其他連結的節點，進而引發更多的網路節點被激發，使得整個網路結構呈現活化的狀態。在此過程的最終階段，相鄰且具有強度連結的節點，它們所呈現出來的重要程度會趨近於相似。

iSpreadRank 共可分為三個階段，依序是：(1)初始化 (Initialization)、(2)推論 (Inference) 及(3)預測 (Prediction)。初始化步驟的目的是將語句關係網路轉化為 Sentence-by-Sentence 的矩陣，以利推論步驟的運作。矩陣轉化如 Eq. (14) 所示：

$$A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{bmatrix} \quad (14)$$

其中， $a_{i,i} = 0.0$ ， $a_{i,j} = w(e_{i,j})$ 。

iSpreadRank 以 Leaky Capacitor Model (Anderson 1983) 進行推論步驟的實作。假設 $X(0)$ 為一 $n \times 1$ 的向量，向量中的任一元素 x_i 是語句 s_i 的初始向心性，藉由 Eq. (15) 得到所有語句節點在第 t 時間點的 iSpreadRank Centrality，即是 $X(t)$ ：

$$X(t) = X(0) + MX(t-1), \quad M = \sigma R^T \quad (15)$$

此處， σ 為一擴散因子 (Spreading Factor)，用來控制節點間投票的有效性。考慮節點 A 投了一票給節點 B ，若 σ 為 0.5 的話，則表示 A 投給 B 的那一票僅具有一半的效力，即相當於 0.5 張票。實作時將 $X(0)$ 中的元素初始值設定為 1， σ 設定成 0.7。 R 是一隨機矩陣 (Stochastic Matrix)，定義成 Eq. (16)：

$$R = \begin{bmatrix} r_{1,1} & \cdots & r_{1,n} \\ \vdots & \ddots & \vdots \\ r_{n,1} & \cdots & r_{n,n} \end{bmatrix} \quad (16)$$

其中， $r_{i,j} = \frac{a_{i,j}}{\sum_k a_{i,k}}$ 。由 Eq. (16)可知：矩陣 R 的每列數值總和皆是 1；亦即，列 i 具有 $\sum_j r_{i,j} = 1$ 的特性。考慮離散語句關係網路 G' 時，Eq. (16)的功用是用來平均分配每個語句節點的影響力給所有正向連結所指向的節點；而以連續語句關係網路 G 來說，其分配則是按照連結權重與相連節點之平均關聯強度比例來計算。

在推論的過程之中，iSpreadRank 由 Power Method 經過迭代的方式模擬該網路模型的擴散激發機制，用以求得網路活化狀態趨於穩定時最後狀態。迭代過程的停止時機可由事先給定的迭代數目決定，或是由網路節點的活化狀態是否已經趨於穩定判斷。實務上適當的迭代數目通常不易決定，必須由事先的實驗或機器學習過程得到或是由領域專家來給定。本研究採用第二種策略將推論機制的停止條件設定成第 t 時間與第 $t-1$ 時間點的網路節點活化值差異總和是否小到可以忽略，而其判斷網路活化狀態穩定而停止推論的條件設定如下：

$$\sum_i |X_i(t) - X_i(t-1)| \leq \varepsilon \quad (17)$$

實作時將 Eq. (17)中的 ε 設定為 0.0001。

最後預測的階段中，當推論過程停止時，此時網路的活化狀態趨於穩定時的最後狀態，而 $X(t)$ 向量即被視為語句節點經過活化之後得到的預測向心性。依據 $X(t)$ 的向量元素值大小排序，可以定義語句的排名序列作為之後摘要生成的參考。

特別一提的是，當網路中的語句節點不具有任何連結到其他節點時，這類節點在推論過程中並不會被激發活化；換句話說，這類節點最後所得到的向心性與其 $X(0)$ 中所定義的初始值相同。

四、語句排序與摘要生成

語句排序模組根據語句向心性的數值將語句進行排序；排序的原則是由權重值較高的語句排到權重值較低的語句，最後形成一個語句排序序列。接著，依序進行語句的擷取得到摘要的語句集合。

摘要生成模組包括語句擷取和摘要組成。在語句擷取的過程中，考量語句取自多篇主題相關的文件而易將重複資訊 (Redundancy) 納入到摘要中，因而必須過濾多餘的資訊，以避免在有限的摘要長度內擷取到含括重複訊息的語句。過去的文獻中提出懲罰因子 (Penalty) 的想法，例如：Goldstein 等 (2000) 的 MMR-MD (Maximal Marginal Relevance Multi-Document)。這種作法在擷取某個語句到摘要語句集合的當下，須先考慮其與現有摘要語句的相似度，接著依此相似度定義懲

罰因子的權重，藉此調整該語句的重要性，達到降低其被誤判成摘要語句之可能性的目的。

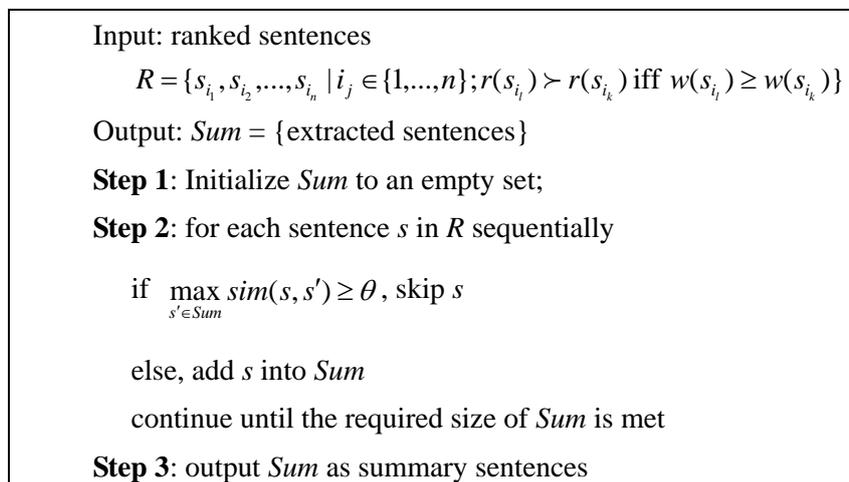


圖 7：引用 CSIS (Radev et al. 2004) 而設計的摘要語句擷取演算法

本研究引用 CSIS (Cross-Sentence Information Sub-sumption) (Radev et al. 2004) 來克服資訊重複性的問題：假設語句 a 與語句 b 的資訊重複時，若且唯若 a 擁有的資訊部份或全部含括於語句 b 所要表達的資訊中。實作時計算兩兩語句的相似度以決定它們是否資訊被彼此所包含，進而設計如圖 7 所示之摘要語句擷取演算法；其中， R 是語句的排序序列， Sum 是摘要語句的集合， θ 是判斷語句是否資訊重複的臨界值，依照實際需要設為 0.7。簡而言之，圖 7 依序選取摘要語句，然在選取的同時需評估目前的候選語句與先前已被擷取之摘要語句的相似度；若相似度大於事先定義的臨界值時，表示該候選語句具有高度的重複資訊，此時應略過避免擷取到該候選語句。

在摘要組成的最後階段，將挑選出來的摘要語句依照其在原始新聞文件中的順序及所屬新聞文件的發佈日期進行排序以形成最後的輸出摘要。此處，語句排序的目的是將語句依據內容的一致性 (Cohension) 和連貫性 (Coherence) 進行重新排列，藉以提供讀者語意連貫性高且適合閱讀的摘要內容。

肆、實驗結果

摘要系統的優劣取決於其產出摘要的品質；評估方法通常可為內在評估 (Intrinsic) 和外在評估 (Extrinsic) 兩類 (Mani & Maybury 1999)。內在評估以人工摘要為理想的摘要，比較機器摘要與人工摘要的異同；例如，計算兩者在字詞

層面相似程度的準確率 (Precision) 或召回率 (Recall)。實務上須由比較不同專家的摘要，再加以平均取得較為客觀的評估結果，藉此降低專家的主觀意見在人工摘要過程中引起的偏差。另外，亦可直接由人工閱讀主觀判定機器摘要的好壞；常用的評估準則包含：主題涵蓋度、內容可讀性、連貫性、一致性、組織性及摘要長度等。相較來說，外在評估則是使用機器摘要取代原始的全文，以評估機器摘要對於其他應用或決策的影響。例如：以文件分類來說，比較使用摘要進行分類與使用原文進行分類是否得到相近的分類準確度。換言之，若機器摘要的品質優良足以取代原文時，則在其他應用上預期也會得到不錯的成效。

本研究採用的評估方法屬於內在評估。以下依序介紹實驗階段所使用的測試資料、評估標準和實驗結果。

一、測試資料集

DUC (Document Understanding Conference)⁹ 是美國 NIST 國家標準與技術研究所發起，自 2001 年至 2007 年間考量不同的研究目的提供標準的測試資料集，以利文件摘要的發展及建立實驗評估的流程與標準。本研究採用 DUC 2004 資料集來驗證所提出之摘要方法的可行性和成效¹⁰；DUC 2004 共有五項評比：(1) Task 1: Very short single-document summaries；(2) Task 2: Short multi-document summaries focused by TDT events；(3) Task 3: Very short cross-lingual single-document summaries；(4) Task 4: Short cross-lingual multi-document summaries focused by TDT events；及(5) Task 5: Short summaries focused by question。考量本文所提之摘要方法的適用性，實驗時遵照 Task 2 的規範來進行¹¹。

Task 2 的資料集共有 50 個英文新聞文件群，每個文件群約有 10 篇取自於 Associated Press 與 New York Times 兩家報社的新聞，且同一個文件群的新聞皆是報導相同的新聞事件；此點符合圖 3 之新聞事件摘要架構中輸入為多篇主題相關的新聞。Task 2 要求對 50 個新聞文件群個別產生約 665 bytes 長度的短摘要。有關人工摘要的產生，每個文件群都至少由 4 位專家閱讀後，針對該文件群所描述的新聞事件，透過人工撰寫產出約 665 bytes 的摘要作為評估時的參考¹²。特別一提的是，人工摘要可能因為專家對於新聞事件所屬領域的不熟悉而造成摘要內容的偏差；因此，Task 2 的 50 個新聞文件群乃是由 8 位 NIST 的專家從 TDT (Topic

9 <http://duc.nist.gov/>.

10 DUC 提供 DUC 2001 至 DUC 2007，總計有 7 個標準資料集。考量本研究的範疇屬於多文件摘要技術的研發，因而選用 DUC 2004 資料集進行實驗。

11 DUC 2004 的 Task 2 規範，詳請參考 <http://duc.nist.gov/duc2004/tasks.html>。

12 NIST 對於每個新聞文件群要求專家撰寫長度約 665 個字元 (即，665 bytes) 的摘要；摘要內容由專家對於新聞事件的解讀自由表述。有關摘要撰寫的規範，詳請參考 <http://duc.nist.gov/duc2004/t1.2.summarization.instructions>。而 665 個字元的摘要長度限制乃是參考 DUC 2003 的平均人工摘要長度決定。

Detection & Tracking) 資料集中依所感興趣的新聞事件來挑選，且每個文件群至少由 4 位專家撰寫摘要，並以評比結果的平均值來降低摘要偏差影響的可能。

在資料的特性方面，每個文件群平均有 266.86 個語句，每篇文章平均有 26.69 個語句。在字數統計方面，每個文件群平均有 6,726.88 個字，每篇文章平均有 672.69 個字，而每個語句平均有 25.21 個字。在人工摘要的部分，每篇摘要平均有 3.88 個語句，換算平均約 106.48 個字；而平均每個摘要語句具有 28.21 個字。根據前述的統計說明，換算得到的摘要壓縮比約為 1.5%。

二、評估標準

ROUGE(Recall-Oriented Understudy for Gisting Evaluation)(Lin & Hovy 2003) 以專家摘要為參考標準，計算專家與機器產生的摘要所共同包含的字詞比例。考量字詞單元與評估範圍框定之差異，ROUGE 定義的評量標準有 ROUGE-1, ROUGE-2, ..., ROUGE-N, ROUGE-L, ROUGE-WL；其中，ROUGE-N 以 N 個字為單位計算召回率。例如，ROUGE-2 是以二字詞 (Bi-gram) 為字詞單位得到的召回率；ROUGE-L 是以最長相同序列 (Longest Common String) 為單位計算的召回率；ROUGE-WL 則是加權過後的 ROUGE-L。過去研究已經證實 ROUGE-1 的評估標準接近於專家評比摘要品質的標準 (Lin & Hovy 2003)；因此，本文中所列之實驗數據僅針對 ROUGE-1 進行討論。

三、實驗參數設定

實驗參數設定的策略主要依據相關文獻提供參考，同時藉由前置實驗的分析得到；詳如表 1 所列。前置實驗的流程是隨機從 DUC 2004 資料集中挑選 10 個文件群，接著使用不同參數值進行測試；重複前述步驟共 5 次，計算在不同參數值設定下的 ROUGE-1 值平均，並以擁有最高 ROUGE-1 的參數值作為正式實驗時的設定。

表 1：實驗參數的設定

	參數設定	設定參考
Eq. (4)、Eq. (5)	$\alpha = 0.1$	Erkan & Radev 2004
Eq. (13)	$d = 0.85$	Brin & Page 1998
Eq. (15)	$\sigma = 0.7$	前置實驗決定
Eq. (17)	$\varepsilon = 0.0001$	Haykin 2008
圖 7	$\theta = 0.7$	前置實驗決定

四、評估結果

(一) 不同語句向心性之摘要方法的可行性分析

表 2 與表 3 列出基於連續權重語句關係網路 G' 及離散權重語句關係網路 G'' 時，使用不同語句向心性計算之摘要方法的 ROUGE 數值。Rand. Baseline 與 NIST Baseline 是用來做為比較的參考基準：Rand. Baseline 賦予每個語句隨機的權重來擷取語句產生摘要，其各項 ROUGE 值為 10 次隨機結果的平均；NIST Baseline 是 DUC 2004 所提供的基準，其摘要取自於文件群中最新的那篇新聞之前面 665 bytes 的內容。

表 2 顯示 iSpreadRank 有最好的實驗結果；除了 ROUGE-L 之外，其在各種 ROUGE 評估標準的表現都得到最高的分數。考慮非遞迴式向心性計算時，即 Degree 與 Normalized Similarity-based Degree，兩者在 ROUGE-1 的指標下表現相近；Normalized Similarity-based Degree 比 Degree 僅有相對提升 1.78%¹³。考慮遞迴式向心性計算時，即 HITS、PageRank 與 iSpreadRank，由表發現 iSpreadRank 與 PageRank 表現較佳；在 ROUGE-1 的指標下，iSpreadRank 比 PageRank 相對提高出 1.18%，但比 HITS 相對提升 16.3%。在此三種演算法之中，HITS 的表現最差，甚至比 Rand. Baseline 的結果糟糕；究其原因可能是 HITS 同時考量節點的 In-Degree 與 Out-Degree，而本文的語句關係網路將連結視為雙向連結，無法區別出正向與反向連結，因而導致 HITS 得到最差的數據。

表 2：基於連續權重語句關係網路 G' ，使用不同語句向心性計算之摘要方法的 ROUGE 數值

Models	ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-4	ROUGE-L	ROUGE-W1.2
Rand. baseline	0.31549	0.04634	0.01031	0.00368	0.33878	0.11459
NIST baseline	0.32419	0.06411	0.01992	0.00671	0.34590	0.11859
Degree_ G'	0.35296	0.07090	0.02102	0.00833	0.35826	0.12224
Normalized Similarity-based Degree_ G'	0.35926	0.07010	0.01993	0.00697	0.37015	0.12582
HITS_ G'	0.31142	0.05405	0.01565	0.00691	0.33203	0.11327
PageRank_ G'	0.35796	0.07315	0.02089	0.00742	0.35962	0.12273
iSpreadRank_ G'	0.36218	0.07961	0.02491	0.01065	0.36989	0.12657

¹³ b 比 a 的相對提升比例，計算方式： $(b-a)/a*100\%$ 。

表 3：基於離散權重語句關係網路 G'' ，使用不同語句向心性計算之摘要方法的 ROUGE 數值

Models	ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-4	ROUGE-L	ROUGE-W1.2
Rand. baseline	0.31549	0.04634	0.01031	0.00368	0.33878	0.11459
NIST baseline	0.32419	0.06411	0.01992	0.00671	0.34590	0.11859
Degree_ G''	0.33743	0.06760	0.01990	0.00850	0.34763	0.11892
Normalized Similarity-based Degree_ G''	0.34976	0.06391	0.01724	0.00559	0.35814	0.12185
HITS_ G''	0.34485	0.06597	0.01880	0.00697	0.34972	0.11935
PageRank_ G''	0.35366	0.07499	0.02288	0.00978	0.35977	0.12338
iSpreadRank_ G''	0.36047	0.07385	0.02141	0.00778	0.36444	0.12424

由表 3 可知，PageRank 在 ROUGE-2、ROUGE-3 和 ROUGE-4 的評估標準得到最高的分數；而 iSpreadRank 在 ROUGE-1、ROUGE-L 及 ROUGE-W1.2 的評估標準得到最高的分數。考慮非遞迴式向心性計算時，在 ROUGE-1 的指標下，Normalized Similarity-based Degree 比 Degree 相對提高高出 3.65%。考慮遞迴式向心性計算時，實驗發現在 ROUGE-1 的指標下 iSpreadRank 的表現最好，比 PageRank 相對提高高出 1.93%，比 HITS 相對提高高出 4.53%。而在此三種演算法之中，HITS 的表現最差，PageRank 的表現居中。同樣的，本研究將語句關係網路中的連結視為雙向連結，無法確實反應出正向與反向連結的差別，因而導致 HITS 得到最差的數據。

接著，比較基於連續權重語句關係網路 G' 與離散權重語句關係網路 G'' 對於語句向心性計算的影響，實驗結果顯示：Degree_ G' 比 Degree_ G'' 相對提高高出 4.6%；Normalized Similarity-based Degree_ G' 比 Normalized Similarity-based Degree_ G'' 相對提高高出 2.72%；PageRank_ G' 比 PageRank_ G'' 相對提高高出 1.22%；iSpreadRank_ G' 比 iSpreadRank_ G'' 相對提高高出 0.47%。但是，HITS_ G'' 卻比 HITS_ G' 相對提高高出 10.73%。由前述可知，除了 HITS 之外，基於連續權重語句關係網路 G' 計算語句向心性的摘要方法都比基於離散權重語句關係網路 G'' 的方法表現來得佳。

綜合前述的實驗結果：在 ROUGE-1 指標下，各種向心性計算的表現依序是：iSpreadRank > Normalized Similarity-based Degree > PageRank > Degree > HITS。

最後，表 4 是 DUC 2004 的部分比賽結果；SYSID 是不同的參賽系統，字母 A-H 代表專家摘要與其他專家摘要的比較結果，數字代表當年度的摘要系統。由表

得知，本文所提出的摘要方法中，Degree_ G' 、Normalized Similarity-based Degree_ G' 、PageRank_ G' 、HITS_ G'' ，與 iSpreadRank_ G' 都有不錯的結果，其表現比 DUC 2004 的 Median Machine 及機器的 ROUGE-1 平均值來得佳。由此說明本研究利用語句向心性來評估語句摘要代表性以排序語句的方法確實可行。

表 4：DUC 2004 的部分比賽結果與本文所提的摘要方法效能比較

SYSID	ROUGE-1	95% Confidence Interval
H	0.41828	[0.40193, 0.43463]
F	0.41246	[0.39161, 0.43331]
E	0.41038	[0.38817, 0.43259]
D	0.40594	[0.38700, 0.42488]
B	0.40428	[0.37946, 0.42910]
A	0.39325	[0.37218, 0.41432]
C	0.39039	[0.37149, 0.40929]
G	0.38902	[0.36793, 0.41011]
Mean of Human Assessors	0.40300	[0.38247, 0.42353]
2 (NIST Baseline) (Rank: 25/35)	0.32419	[0.30922, 0.33916]
Best Machine (SYSID=65)	0.38224	[0.36941, 0.39507]
Median Machine (SYSID=138)	0.34299	[0.32805, 0.35793]
Worst Machine (SYSID=111)	0.24190	[0.23038, 0.25342]
Mean of Machines	0.33683	[0.32258, 0.35108]
Degree_ G'	0.35296	[0.33827, 0.36765]
Normalized Similarity-based Degree_ G'	0.35926	[0.34674, 0.37178]
HITS_ G''	0.34485	[0.32961, 0.36009]
PageRank_ G'	0.35796	[0.34489, 0.37103]
iSpreadRank_ G'	0.36218	[0.34611, 0.37825]

(二) 整合語句向心性和語句位置特徵之摘要方法的效能比較

Erkan 與 Radev (2004) 應用 PageRank 計算語句向心性，進而提出 LexRank 作為多文件摘要的模型，用以標示語句為摘要或非摘要語句；該研究除了評估語句向心性之外，同時結合語句位置特徵得到不錯的摘要成效。本節參考他們的實驗架構，將語句向心性視為特定的語句特徵，並以線性組合整合語句向心性和語句位置特徵作為衡量語句重要性的依據。實作時固定語句位置特徵的權重為 1.0，

接著考量不同的語句向心性權重條件來整合；依據 Erkan 與 Radev (2004) 的實驗規範，語句向心性的權重設定為 0.5、1.0、1.5、2.0、2.5、3.0、5.0 和 10.0。

表 5 列出整合語句向心性和語句位置特徵之摘要方法的 ROUGE-1 值，其中：Avg、Max 與 Min 表示語句向心性在不同權重條件下的平均值、最大值及最小值；POS 指的是語句位置特徵；Baseline I 和 Baseline II 是參考基準，相關數據節錄於 Erkan 與 Radev (2004)，依序為 LexRank 在基於離散及連續權重語句關係網路時的評比結果。由表得知，在基於離散權重語句關係網路 G'' 的架構下，POS+PageRank_ G'' 與 POS+iSpreadRank_ G'' 的平均表現都高於 Baseline I，相對提升高出 0.13% 和 1.24%。而在基於連續權重語句關係網路 G' 的架構下，POS+PageRank_ G' 的平均表現約等同於 Baseline II；僅有 POS+iSpreadRank_ G' 的平均表現優於 Baseline II，相對提升高出 0.61%。

接著，在整合語句位置特徵的設定下，比較基於連續權重語句關係網路 G' 與離散權重語句關係網路 G'' 之語句向心性的效能差異，表 5 顯示在 ROUGE-1 指標下，各種向心性的表現依序是：iSpreadRank > PageRank > Normalized Similarity-based Degree > Degree > HITS。比較表 2、表 3 和表 5 發現，在將語句向心性視為特定的語句特徵，同時整合語句位置特徵以評估語句重要性的摘要架構下，其表現相對於沒有整合語句位置特徵的結果擁有不錯的增進。例如：POS+iSpreadRank_ G' 比 iSpreadRank_ G' 提升高出 4.40%；POS+iSpreadRank_ G'' 比 iSpreadRank_ G'' 提升高出 4.92%。

表 5：整合語句向心性和語句位置特徵之摘要方法的 ROUGE-1 數值

Models	Avg.	Max.	Min.
Baseline I: LexRank	0.3736	0.3808	0.3646
Baseline II: Cont. LexRank	0.3758	0.3826	0.3617
POS+Degree_ G'	0.37251	0.38030	0.36109
POS+Normalized Similarity-based Degree_ G'	0.37528	0.38191	0.36354
POS+HITS_ G'	0.33810	0.37631	0.32077
POS+PageRank_ G'	0.37574	0.38008	0.37143
POS+iSpreadRank_ G'	0.37811	0.38634	0.37325
POS+Degree_ G''	0.36279	0.37825	0.34546
POS+Normalized Similarity-based Degree_ G''	0.37061	0.37747	0.36192
POS+HITS_ G''	0.35648	0.37834	0.34754
POS+PageRank_ G''	0.37407	0.38241	0.36216
POS+iSpreadRank_ G''	0.37822	0.38113	0.37174

表 6：整合語句向心性和語句位置特徵之摘要方法與其他圖形結構相關方法的效能比較

Methods	ROUGE-1	95% Confidence Interval
1.0POS+1.0Degree_G'	0.37646	[0.36252, 0.39040]
1.0POS+1.0Degree_G''	0.37675	[0.36315, 0.39035]
1.0POS+1.0Normalized Similarity-based Degree_G'	0.38191	[0.36993, 0.39389]
1.0POS+1.0Normalized Similarity-based Degree_G''	0.37195	[0.35762, 0.38628]
1.0POS+1.0HITS_G'	0.35519	[0.33759, 0.37279]
1.0POS+1.0HITS_G''	0.36467	[0.34794, 0.38140]
1.0POS+1.0PageRank_G'	0.37865	[0.36471, 0.39259]
1.0POS+1.0PageRank_G''	0.38241	[0.36969, 0.39513]
1.0POS+1.0iSpreadRank_G'	0.38634	[0.37270, 0.39998]
1.0POS+1.0iSpreadRank_G''	0.37954	[0.36579, 0.39329]
S65@DUC2004 (Rank: 1/35)	0.38224	[0.36941, 0.39507]
S104@DUC2004 (Rank: 2/35)	0.37443	[0.36354, 0.38532]
S35@DUC2004 (Rank: 3/35)	0.37430	[0.36121, 0.38739]
LexRank (Erkan & Radev 2004)	0.3736	N/A
Cont. LexRank (Erkan & Radev 2004)	0.3758	N/A
Wan & Yang (2006)	0.41102	N/A
Grasshopper (Zhu et al., 2007)	0.3755	[0.3622, 0.3888]
Cluster-HITS (Wang & Yang 2008) [Re-evaluated in Cai & Li 2013]	0.36463	N/A
Graph-Sum (Boudin et al. 2011)	0.38052	N/A
Canhasi & Kononenko 2011	0.379	[0.361, 0.389]
G-FLOW (Christensen et al. 2013)	0.3733	N/A
Cai & Li 2013	0.37475	N/A

表 6 是整合語句向心性和語句位置特徵之摘要方法與其他圖形結構相關方法的效能比較。前 10 項是基於連續和離散權重之語句關係網路，以相同權重整合語句向心性和語句位置特徵的摘要方法，例如：1.0POS+1.0Degree_G 表示線性組合中語句位置特徵權重與語句向心性權重都設定 1.0；S65、S104 和 S35 是 DUC 2004 當年度排名前 3 名的參賽系統；其餘項目則是摘錄其他圖形結構相關方法文獻的

評比數據。

由表可知，相較於 DUC 2004 的參賽系統，本節中整合語句向心性和語句位置特徵的摘要方法，除了 POS+Normalized Similarity-based Degree_G'、POS+HITS_G' 及 POS+HITS_G'' 之外，絕大多數的評估結果都介於 S65 和 S104；POS+PageRank_G'' 與 POS+iSpreadRank_G' 的表現甚至優於 S65，提升的比例依序是 0.04% 和 1.07%。而相較於其他圖形結構相關的摘要技術而言，本研究整合語句向心性和語句位置特徵的摘要方法同樣擁有不錯的效能表現。

整體來說，比較相關文獻同時整合圖形結構分析與其他技術，本文所提方法在單純使用語句向心性衡量語句代表性的設定下，其效能可能沒有比其他研究所提的方法來得好；但若整合語句向心性和語句位置特徵時，其表現則是優於其他圖形結構相關的方法。未來方向可朝著整合其他類型的語句特徵繼續研究，以探討並找出適當的特徵組合架構；或是整合不同類型的文件分析技術發展更好的摘要架構。

伍、討論

本文所提出的「應用語句關係網路計算語句向心性之新聞事件摘要」具有以下特性：(1)此方法屬於非監督式 (Unsupervised)，有別於監督式機器學習需要事先收集訓練資料以進行分類模型的學習；(2)此方法並未考慮到領域知識，亦未針對語言的特性進行處理，可應用到其他領域中而不必大幅度調整摘要架構與模組設定；(3)此方法的設計已模組化，具有彈性可依需要進行擴充或進行個別模組的替換與修正。但所謂一刀兩刃，前面所提及的特性背後卻也隱藏著先天的限制。例如，若能加入領域知識的考量、文件結構的分析，或是語言特性的處理，相信有助於摘要系統效能及摘要品質的提升。除此之外，本研究以摘錄來產生多文件摘要，其缺點是直接從個別的文件中擷取出摘要語句，難以保證最後輸出的摘要內容組織之連貫性及一致性上具有不錯的水準。

本研究的主要貢獻，包含：(1)延伸單文件摘要適用之主題關係地圖 (Salton et al. 1997)，提出適用於多文件摘要的語句關係網路作為新聞事件的主題模型；同時考量連結關係的權重，定義連續權重及離散權重兩種語句關係網路模型；(2)基於語句關係網路模型，導入圖形結構理論用來評估節點向心性的分析法則，提出五種語句向心性計算的演算法；同時利用語句向心性來決定語句的摘要代表性，藉以達到排序語句的目的；(3)使用 DUC 2004 資料集驗證本文所提出之摘要方法的可行性和成效，同時探討整合語句向心性和語句位置特徵作為語句重要性評估之摘要方法效能，相關經驗可供其他研究學者參考比較。

陸、結論

本文提出「應用語句關係網路計算語句向心性之新聞事件摘要」方法：首先，延伸 Salton 等（1997）所提出的主題關係地圖，將語句視為網路中的節點，計算兩兩語句間的關聯程度以決定語句間是否存在有語意連結，進而構成語句關係網路作為新聞事件的主題模型。另外，考量到語句關係網路中的連結具有權重，本研究共提出「連續權重語句關係網路 (G')」與「離散權重語句關係網路 (G'')」兩種不同的網路模型。接著，引用圖形結構理論中用來描述與量化網路結構的法則，衡量節點在網路的重要性或對其他相連節點的影響性，並以節點的向心性來決定語句的重要程度，用以作為語句的摘要代表性；同時設計五種節點向心性的計算方法，包含：(1) Degree Centrality、(2) Normalized Similarity-based Degree Centrality、(3) HITS Centrality、(4) PageRank Centrality，以及(5) iSpreadRank Centrality。最後，基於節點的向心性將所有語句進行排序，再依此語句排序序列作為摘要生成之語句擷取的根據。在語句擷取的過程中，導入 CSIS (Radev et al. 2004) 過濾重複資訊，並在固定摘要長度的條件下擷取適當數量的語句，最後考量語意的連貫性重組輸出生成的摘要文本。

實驗使用 DUC 2004 資料集來驗證本文所提出之摘要方法的可行性及成效。摘要品質的優劣則以 ROUGE 來進行評估，並採用 ROUGE-1 指標作為評比的基準。實驗結果顯示：在 ROUGE-1 指標下，各種向心性計算的表現依序是：iSpreadRank > Normalized Similarity-based Degree > PageRank > Degree > HITS。除了 HITS 之外，基於連續權重語句關係網路 G' 計算語句向心性之摘要方法皆比基於離散權重語句關係網路 G'' 的表現來得佳。比較 DUC 2004 當年度的評比結果時，本文所提出的摘要方法中，包含 Degree_ G' 、Normalized Similarity-based Degree_ G' 、PageRank_ G' 、HITS_ G'' ，以及 iSpreadRank_ G' 都有不錯的效能，其表現都比 DUC 2004 的 Median Machine 和所有機器的 ROUGE-1 平均值來得好。整體來說，實驗發現本研究利用語句向心性來評估語句摘要代表性以排序語句的方法確實是可行的方式。

實驗另外發現在將語句向心性視為特定的語句特徵，同時整合語句位置特徵以評估語句重要性的摘要架構下，其表現相較於沒有整合語句位置特徵的結果擁有不錯的增進；除了 POS+Normalized Similarity-based Degree_ G'' 、POS+HITS_ G' 及 POS+HITS_ G'' 之外，絕大多數的評估結果對照 DUC 2004 的機器來說，排名都介於第一及第二名之間。另比較其他圖形結構相關的摘要技術，本文中整合語句向心性和語句位置特徵的摘要方法同樣擁有不錯的效能表現。

特別說明的是：本文以新聞事件摘要為例說明，實驗亦使用新聞資料集進行方法成效的驗證；然而，本研究所提出的摘要方法同時適用於其他主題相關之多

文件的輸入。

未來可能的研究方向有：(1)語句關係網路中的連結實際上並未真正考慮到語句節點的語意關聯，建議藉由字詞的語義解析來定義語句關聯，用以將關鍵字詞層面的相似度提升到語意層面的分析；(2)社群網路分析 (Social Network Analysis) 的相關研究已研發出多種節點重要性的評估方法，例如：近距向心性 & 參與向心性 etc.，建議引用這些研究成果來改良本文所提出的語句向心性計算；(3)目前在語句向心性分析的方法中，不論是 HITS、PageRank 或 iSpreadRank 都假設語句節點的初始權重值為 1.0，若能借助傳統文件摘要的語句特徵分析，相信可得到較佳的語句節點權重初始值；(4)實驗發現整合語句向心性和語句位置特徵的方法比單純使用語句向心性的方法表現好，建議整合其他類型的語句特徵繼續研究，以探討並找出適當的特徵組合架構；(5)摘錄式摘要的缺點是無法保證其內容在語意連貫性及可讀性上有不錯的品質，建議使用語句生成技術重新編寫產生概述式摘要。

誌謝

本研究由國科會計畫 (NSC99-2221-E-259-018-MY3: 以圖形網路模型為基礎之文件摘要方法設計與研究) 提供經費補助，特此致謝。

參考文獻

- 陳光華 (1998), 『新資訊時代的啟發性資訊服務』, *21 世紀資訊科學與技術的展望學術研討會論文集*, 桃園, 臺灣, 頁 195-208。
- 葉鎮源 (民 91), 『文件自動化摘要方法之研究及其在中文文件的應用』, 未出版碩士論文, 國立交通大學資訊科學系, 新竹。
- 謝仰哲 (民 97), 『國中生友誼與學習諮詢網路之社會網路分析』, 未出版碩士論文, 國立臺灣師範大學資訊教育學系, 臺北。
- Aliguliyev, R.M. (2010), 'Clustering techniques and discrete particle swarm optimization algorithm for multi-document summarization', *Computer Intelligence*, Vol. 26, No. 4, pp. 420-448.
- Anderson, J.R. (1983), 'A spreading activation theory of memory', *Journal of Verbal Learning and Verbal Behavior*, Vol. 22, No. 3, pp. 261-295.
- Barzilay, R. and Elhadad, M. (1997), 'Using lexical chains for text summarization', *Proceedings of the ACL/EACL'97 Workshop on Intelligent Scalable Text Summarization*, Madrid, Spain, July 11, pp. 10-17.
- Barzilay, R., McKeown, K.R. and Elhadad, M. (1999), 'Information fusion in the context of multi-document summarization', *Proceedings of the 37th Annual*

- Meeting of the Association for Computational Linguistics (ACL 1999)*, College Park, MD, June 20-26, pp. 550-557.
- Boudin, F., Huet, S. and Torres-Moreno, J.M. (2011), 'A graph-based approach to cross-language multi-document summarization', *Polibits*, Vol. 43, pp. 113-118.
- Brin, S. and Page, L. (1998), 'The anatomy of a large-scale hypertextual Web search engine', *Computer Networks and ISDN Systems*, Vol. 30, No. 1-7, pp. 107-117.
- Cai, X. and Li, W. (2013), 'Ranking through clustering: an integrated approach to multi-document summarization', *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 21, No. 7, pp. 1424-1433.
- Canhási, E. and Kononenko, I. (2011), 'Semantic role frames graph-based multidocument summarization', *Proceedings of the 14th International Multiconference on Information Society*, Ljubljana, Slovenia, October 10-14, pp. 113-116.
- Carrington, P.J., Scott, J. and Wasserman, S. (2005), *Models and methods in social network analysis*, Cambridge University Press, New York, NY.
- Christensen, J., Mausam, Soderland, S. and Etzioni, O. (2013), 'Towards coherent multi-document summarization', *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Atlanta, GA, June 9-14, pp. 1163-1173.
- Collins, A.M. and Loftus, E.F. (1975), 'A spreading-activation theory of semantic processing', *Psychological Review*, Vol. 82, No. 6, pp. 407-428.
- Daniel, N., Radev, D. and Allison, T. (2003), 'Sub-event based multidocument summarization', *Proceedings of the HLT-NAACL'03 Workshop on Text Summarization*, Edmonton, Canada, May 30, pp. 9-16.
- Das, D. and Martins, A.F.T. (2007), 'A survey on automatic text summarization', available at <http://www.dipanjandas.com/files/summarization.pdf> (accessed 17 March, 2014).
- Erkan, G. (2006), 'Using biased random walks for focused summarization', *Proceedings of the DUC 2006 Document Understanding Workshop*, Brooklyn, NY, June 8-9.
- Erkan, G. and Radev, D.R. (2004), 'LexRank: graph-based lexical centrality as salience in text summarization', *Journal of Artificial Intelligence Research*, Vol. 22, pp. 457-479.
- Goldstein, J., Mittal, V., Carbonell, J. and Kantrowitz, M. (2000), 'Multi-document summarization by sentence extraction', *Proceedings of the 2000 NAACL-ANLP Workshop on Automatic Summarization*, Seattle, WA, April 30, pp. 40-48.

- Gong, Y. and Liu, X. (2001), 'Generic text summarization using relevance measure and latent semantic analysis', *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, New Orleans, LA, September 9-12, pp. 19-25.
- Gupta, V., Chauhan, P. and Garg, S. (2012), 'An statistical tool for multi-document summarization', *Journal of Scientific and Research Publications*, Vol. 2, No. 5, pp. 1-5.
- Harabagiu, S. and Maiorano, S. (2002), 'Multi-document summarization with GISTexter', *Proceedings of the 3rd LREC Conference (LREC 2002)*, Canary Islands, Spain, May 29-31, pp. 1456-1463.
- Haveliwala, T.H. (2002), 'Topic-sensitive PageRank', *Proceedings of the 11th International Conference on World Wide Web (WWW 2002)*, Honolulu, HI, May 7-11, pp. 517-526.
- Haykin, S.O. (2008), *Neural networks and learning machines*, Prentice Hall, New York, NY.
- Ji, H., Favre, B., Lin, W.P., Gillick, D., Hakkani-Tur, D. and Grishman, R. (2013), 'Open-domain multi-document summarization via information extraction: challenges and prospects', in Poibeau, T., Saggion, H., Piskorski, J. and Yangarber, R. (Eds.), *Multi-source, Multilingual Information Extraction and Summarization*, Springer, Berlin, Heidelberg, Germany, pp. 177-201.
- Kleinberg, J.M. (1999), 'Authoritative sources in a hyperlinked environment', *Journal of the ACM*, Vol. 46, No. 5, pp. 604-632.
- Mani, I. and Bloedorn, E. (1999), 'Summarizing similarities and difference among related documents', *Information Retrieval*, Vol. 1, No. 1-2, pp. 35-67.
- Mani, I. and Maybury, M.T. (1999), *Advances in automatic text summarization*, The MIT Press, Cambridge, MA.
- McDonald, D.M. and Chen, H. (2006), 'Summary in context: searching versus browsing', *ACM Transactions on Information Systems*, Vol. 24, No. 1, pp. 111-141.
- McKeown, K.R., Klavans, J.L., Hatzivassiloglou, V., Barzilay, R. and Eskin, E. (1999), 'Towards multidocument summarization by reformulation: progress and prospects', *Proceedings of the 16th National Conference on Artificial Intelligence (AAAI 1999)*, Orlando, FL, July 18-22, pp. 453-460.
- McKeown, K. and Radev, D.R. (1995), 'Generating summaries of multiple news articles', *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 1995)*, Seattle, WA,

- July 09-13, pp. 74-82.
- Mihalcea, R. (2004), 'Graph-based ranking algorithms for sentence extraction, applied to text summarization', *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, Barcelona, Spain, July 21-26, pp. 170-173.
- Mihalcea, R. and Tarau, P. (2005), 'An algorithm for language independent single and multiple document summarization', *Proceedings of the 2nd International Joint Conference on Natural Language Processing*, Jeju Island, Korea, October 11-13, pp. 19-24.
- Nenkova, A. and McKeown, K. (2012), 'A survey of text summarization techniques', in Aggarwal, C. C. and Zhai, C.X. (Eds.), *Mining Text Data*, Springer, New York, NY, pp. 43-76.
- Li, J. and Li, S. (2013), 'A novel feature-based Bayesian model for query focused multi-document summarization', *Transactions of the Association for Computational Linguistics*, Vol. 1, pp. 89-98.
- Lin, C.Y. and Hovy, E. (2003), 'Automatic evaluation of summaries using N-gram co-occurrence statistics', *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology (HLT-NAACL 2003)*, Edmonton, Canada, May 27 - June 1, pp. 71-78.
- Luhn, H.P. (1958), 'The automatic creation of literature abstracts', *IBM Journal of Research and Development*, Vol. 2, No. 2, pp. 159-165.
- Porter, M.F. (1980), 'An algorithm for suffix stripping', *Program – Automated Library and Information Systems*, Vol. 14, No. 3, pp. 130-137.
- Quillian, M.R. (1968), 'Semantic Memory', in Minsky, M. R. (Ed.), *Semantic Information Processing*, The MIT Press, Cambridge, MA, pp. 227-270.
- Radev, D.R., Jing, H., Styś, M. and Tam, D. (2004), 'Centroid-based summarization of multiple documents', *Information Processing & Management*, Vol. 40, No. 6, pp. 919-938.
- Salton, G. and Buckley, C. (1988), 'Term-weighting approaches in automatic text retrieval', *Information Processing & Management*, Vol. 24, No. 5, pp. 513-523.
- Salton, G., Singhal, A., Mitra, M. and Buckley, C. (1997), 'Automatic text structuring and summarization', *Information Processing & Management*, Vol. 33, No. 2, pp. 193-207.
- Sathish kumar, T. and Sharmila, V. (2013), 'An efficient document summarization using adaptive ranking clustering scheme', *International Journal of Computer*

- Technology & Applications*, Vol. 4, No. 6, pp. 1052-1054.
- Suanmali, L., Salim, N. and Binwahlan, M.S. (2011), 'Genetic algorithm based sentence extraction for text summarization', *Journal of Innovative Computing*, Vol. 1, No. 1, pp. 1-22.
- Yeh, J.Y., Ke, H.R., Yang, W.P. and Meng, I.H. (2005), 'Text summarization using a trainable summarizer and latent semantic analysis', *Information Processing & Management*, Vol. 41, No. 1, pp. 75-95.
- Wan, X. and Yang, J. (2006), 'Improved affinity graph based multi-document summarization', *Proceedings of the 2006 Human Language Technology Conference – North American Chapter of the Association for Computational Linguistics Annual Meeting (HLT-NAACL 2006)*, New York, NY, June 4-9, pp. 181-184.
- Wan, X. and Yang, J. (2008), 'Multi-document summarization using cluster-based link analysis', *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2008)*, Singapore, July 20-24, pp. 299-306.
- Xia, Y., Zhang, Y. and Yao, J. (2011), 'Co-clustering sentences and terms for multi-document summarization', *Proceedings of the 12th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2011)*, Tokyo, Japan, February 20-26, pp. 339-352.
- Zhang, J., Sun, L. and Zhou, Q. (2005), 'A cue-based hub-authority approach for multi-document text summarization', *Proceedings of the 2005 IEEE International Conference on Natural Language Processing and Knowledge Engineering*, Wuhan, China, August 21-23, pp. 642-645.
- Zhu, X., Goldberg, A.B., Van Gael, J. and Andrzejewski, D. (2007), 'Improving diversity in ranking using absorbing random walks', *Proceedings of the 2007 Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2007)*, Rochester, NY, April 22-27, pp. 97-104.