

運用文本挖掘技術分析政府採購法之圍標相關判決案件 Analyzing Bid-Rigging Related Judicial Cases of Government Procurement Law Using Text Mining Techniques

Pei-Zhen Chen
Soochow University
08770022@gm.scu.edu.tw

Hsin-Yun Hsu
Soochow University
crhhaa@gmail.com

Jheng-Long Wu
Soochow University
jlwu@gm.scu.edu.tw

摘要

過往研究多以人工標記判決書資料，並以統計方法來歸納其特徵，故本研究旨在從非結構化數據中萃取關鍵信息，藉由正規表達式進行特徵工程，並找出頻繁出現之項目集，助審計人員更高效準確地選取高風險的政府採購案件。本研究選擇 Apriori 與 LDA 主題分析在資料集中找出違反政府採購法之廠商、採購標案、招標機關之特徵與其頻繁組合。此方法可標示具圍標案件特徵的採購案，並歸納出涉及違反政府採購案常見特徵之頻繁項集，研究結果發現廠商為有限公司，且登記資本額介於 100 萬元至 1,000 萬元，參與最低標金額者出現機率為 0.62，是為較頻繁項及；土木工程承包、建築業務和地方政府合作者距離相近，此些發現在有限審計人力下可成為挑選投標文件的參考之一。

Abstract

Most previous studies have manually labeled judgment data and used statistical methods to summarize their characteristics. Therefore, this study aims to extract key information from unstructured data, perform feature engineering through regular expressions, and find frequently occurring item sets, helping auditors select high-risk government procurement cases more efficiently and accurately. This study uses Apriori and LDA topic analysis to identify the characteristics and frequent combinations of manufacturers, procurement bids, and bidding agencies that violate government procurement laws in the data set. This method can mark procurement cases with the characteristics of bid-rigging cases and summarize frequent items involving violations of

common characteristics of government procurement cases. The research results found that the manufacturer is a limited company with a registered capital between NT\$1 million and NT\$10 million. The probability of occurrence of those who participate in the lowest bid amount is 0.62, which is a relatively frequent project; civil engineering contracting, construction business and local government partners are close to each other. These findings can become one of the references for selecting bid documents under limited audit manpower.

關鍵字：文本挖掘、政府採購法、關聯分析、主題模型

Keywords: Text Mining, Government Procurement, Apriori, LDA

1 前言

每年政府的重要公共工程與政策計畫皆需透過公平透明的採購程序來實現，其中包括工程、資產採購、客製化財物和勞務服務。根據工會統計數據，於 2016 至 2020 年期間，政府機關平均每年處理超過 19 萬筆的採購案件，涉及金額約 1,599 兆新台幣，如表 1 所示（資料來源：工程會官方網站 <https://www.pcc.gov.tw/>），其採購的規模之大。在政府採購中，常見的犯罪行為之一是圍標罪，然而政府採購法尚未明確定義「圍標」一詞，而是描述其模式。當前政府採購圍標案件多為採購專業人員於審標程序中，根據過往經驗人工找出參與競標廠商的投標文件中所存在的異常與其關聯。

根據政府採購法第 109 條的規定，審計機關可以隨時審查政府機關進行的採購活動。同樣地，審計法及其相關細則也賦予政府審計單位通知行政單位提供採購執行情況的權利，並要求各機關提供詳實的答復。故如何

年度	採購決標件數	採購決標金額
105年	18萬餘件	1兆2,050億餘元
106年	18萬餘件	1兆3,725億餘元
107年	19萬餘件	1兆7,644億餘元
108年	19萬餘件	1兆8,716億餘元
109年	20萬餘件	1兆7,817億餘元

表 1. 政府機關採購決標件數及金額情形

有效地選擇有潛在圍標風險的採購案件、進一步請求有關資料，並在有限的審計時間內確定是否存在圍標行為，一直是政府審計人員不斷努力的目標。

在過去的研究中，針對違反政府採購法的判決案例進行探討，主要是根據一定數量的判決案例並進行人工標記，然後利用統計方法歸納出犯罪的特徵和模式。與過往研究不同的是，本研究的目標是運用自然語言處理(NLP)技術，對各級法院判決書進行分析、提取與圍標相關的特徵，並透過頻繁樣式探勘方法與LDA(Latent Dirichlet Allocation)主題分群來辨識出現機率較高之圍標案件特徵組合，進而提升選擇高風險採購案件的效率。

本研究之貢獻：

- 探勘圍標廠商之共同特徵，以此建立自動監控機制，提醒審計人員多加關注該類廠商之投標文件。
- 探勘發生圍標案件之共同特徵，及計算發生機率，輔助審計人員有效挑選符合高風險形態之政府採購案件，爾後新增之判決書亦可自動納入分析，持續更新圍標案件之特徵。

2 文獻探討

2.1 政府採購之犯罪

政府採購領域中的犯罪行為包括「圍標」，雖然政府採購法未明確使用此詞，但其主要刑事責任在於政府採購法第 87 條中。在法院判決書中，對於「圍標」一詞也有深入的討論。圍標被定義為在開標前，具有競爭關係的一些廠商協商，有些廠商不參與投標，或其提出的價格高於已被秘密選定的廠商。這些被選定的廠商則承諾給予協商的其他廠商相應的報酬，這種行為在表面上是合法的，但實際上卻巧妙地規避了競爭(馮世墩, 2018)。本研究透過文字探勘技術對判決書的非結構化資料進行分析，並進行特徵值轉

換，以找出違反政府採購法第 87 條的圍標相關特徵。其貢獻在於在有限的審計人力下，幫助審計人員更有效地選擇具有圍標特徵的採購案件做進一步稽察，從而促進公正、透明的政府採購流程。

2.2 應用 NLP 技術於中文判決書

針對應用自然語言處理(NLP)在判決書案件分類或分群方面，過去部分學者建立混合型案例式推論系統，運用專家知識改進刑事案件的分類，同時引入自動標記語意段落的方法，實現案件段落的擷取和結構化；亦有運用階層分群方法對民事判決要旨進行分群，並透過計算相似程度和關鍵詞加權提升分群效果(廖鼎銘, 2004; 何君豪, 2007)。在判決因素分析與結果預測方面，黃玉婷(2012)運用規則表示式擷取因子並進行分群，以探勘特定類型案件的量刑因子。此外，亦有研究探討了刑事訴訟裁定書的文本一致性與因素分析(陳政瑜, 2015; 黃詩淳及邵軒磊, 2018)。本研究目的在於運用文字探勘技術對非結構化的判決書資料進行特徵值轉換，針對違反政府採購法的判決書，尋找特定的廠商、採購標案和招標機關之特徵。透過有效的特徵探勘方法，審計人員可以更迅速且準確地選擇高風險的政府採購案件，以促進公正透明的政府採購過程。

2.3 特徵工程

頻繁樣式探勘 (Frequent Pattern Mining)：作為資料探勘技術的一環，能夠在不同的資料集中找出重複關係，並透過關聯規則表示出這些關係(Chee 等, 2012)。陳柏慕(2018)收集了台灣 50 成分股在 2013 至 2015 年間的股市交易資料、經濟指標和個股基本資訊，並運用 Apriori 演算法進行頻繁項目集的探勘，歸納出傳產類股、電子類股和金融類股的法人交易邏輯。頻繁樣式探勘技術被證明在發現重要樣式和提取關聯性方面具有強大的潛力，對於本研究欲有效挑選具有違規特徵的政府採購案件，或可提供有價值的方法和洞見。

正規表達式 (Regular Expression)：正規表達式長期以來一直是擷取資訊的主要方法之一。透過正規表達式，可以成功地擷取特定項目，如電子郵件地址、信用卡號、基因及蛋白質

名稱等。這些項目的共同特點在於，他們能夠關鍵代表性的特徵可以透過正規表達式來表示 (Y. Li 等, 2008)。林筱鈺 (2013) 對智慧財產法院著作權相關的民事案件和刑事附帶民事案件判決書進行了探究。他運用了不同的正規表達式規則，分別查詢判決結果、賠償金額、判決案由、影響損害賠償金額因子等內容，並進一步進行研究分析。正規表達式技術在擷取非結構化資料中的特定模式方面具有強大的應用價值，可以用來發掘有關政府採購法違規特徵的案例，並進一步協助研究的目標。

LDA (Latent Dirichlet Allocation) 主題分群：

LDA 使用貝氏網絡 (Bayesian Network) 結合最大期望演算法 (Expectation-Maximization Algorithm) 與條件機率 (Conditional Probability)，將文章內容的主題萃取出來，進而做到文本模組化、文字分類以及協同過濾 (Collaborative Filtering) (Blei, D. M., et al., 2003)。為一種非監督式的機器學習方法，適用於進行大規模的文本分析，如新聞媒體相關之文本 (DiMaggio, et al., 2013)、社群上之推文或評論 (Calheiros et al., 2017) 亦或是文本的分類與摘要 (Calheiros et al., 2017) 等，皆能夠以 LDA 主題分析來進行文本探勘。本研究使用 LDA 之目的在於對違反政府採購法的判決書中所涉及的廠商、採購標案和招標機關進行模組化，找出距離相近的項集，提供更多近似的項集組合，欲提升審計人員在選擇高風險採購案的準確度。

3 研究方法

本研究欲探討涉及圍標案件的廠商、政府採購標案與招標機關有哪些組合。欲瞭解廠商、政府採購案與招標機關有哪些組合較容易發生圍標，擬將其視為較高風險圍標組合，用以辨識 (或標記) 其他政府採購案，作為政府審計人員選案決策之參考，如圖 1 所示。

3.1 擷取判決書資料

本研究所使用之判決書資料先使用爬蟲程式從司法院法律資料檢索 (<https://www.judicial.gov.tw/tw/np-117-1.html>) 系統擷取「政府採購」之裁判書文本資訊。

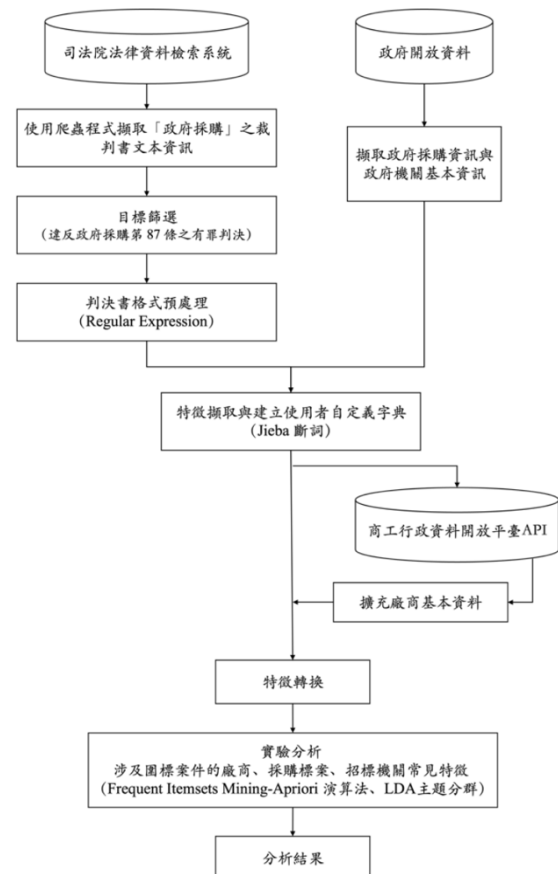


圖 1. 研究流程架構圖

3.2 目標篩選

依刑事訴訟法第 314-1 條規定：「有罪判決之正本，應附記論罪之法條全文。」意即有罪判決在判決書之末，必須揭露所犯罪名的全文法條，本研究聚焦於違反採購法 87 條之案件，經觀察判決書內容，論罪法條全文通常記錄於裁判書文末 (教示條款後)，爰本研究運用正規表達式抓取位於「書記官」後之文字，及揭露之法條，經統計有罪之判決計有 447 件，其中違反採購法 87 條 (圍標) 的裁判書者計有 435 件。

3.3 判決書格式預處理

運用正規表達式 (Regular Expression) 擷取裁判書內提及之廠商名稱、政府採購案名稱。又經觀察裁判書內容，部分機關名稱、採購案名稱 (或稱標案名稱、工程名稱)、廠商名稱資料會以方框繪製字元 (┌、┐、└、┘、├、┤、┆) 繪製類似表格形式於網頁呈現、故

將包含方框繪製字元另作資料清理，再擷取出機關名稱、採購案名稱、廠商名稱。

3.4 擷取政府採購資訊與政府機關基本資訊

至政府電子採購網(<https://web.pcc.gov.tw/pis/>)查詢 96 至 109 年度政府採購決標資訊，。另向經濟部申請並介接使用其資料開放平臺(<https://data.gov.tw/>)。

3.5 特徵擷取與建立使用者自定義字典

- **政府機關名稱字典**：運用 Jieba 斷詞將裁判書文內之機關名稱另成一資料欄位。
- **採購案名稱字典**：運用 Jieba 斷詞將裁判書文內之採購案名稱另成一資料欄位。
- **政府採購標案名稱字典**：將採購案標的名稱另存為文字檔，作為 Jieba 使用者自定義政府採購標案名稱字典。

3.6 擴充廠商基本資料

將裁判書擷取出之廠商名稱，透過商工行政資料開放平臺 API (<https://data.gcis.nat.gov.tw/main/index>) 取得公司、商號的登記資料，共獲得廠商的商工登記基本資料。

3.7 特徵轉換

為降低數值型及字串等資料複雜程度，轉換資料型態為類別型，相關特徵轉換，如表 2 所示。據政機關代碼資料集中，每個機關名稱都有其相對應之機關層級，共 5 類，將招標機關轉換為招標機關層級；另將機關地址擷取出市縣名稱，共為 22 市縣。次依採購法意旨，將政府採購標的分為工程、財物、勞務等 3 類，再以金額大小區分採購金額級距，分別為巨額採購、查核金額、公告金額、未達公告金額、小額採購等 5 類，如表 3 所示。

3.8 實驗分析

本研究欲探討涉及圍標案件中之有罪判決的廠商、採購標案、招標機關常見特徵，並其進行四項分析：違反採購法 87 條（圍標）的廠商有無共同特徵、發生圍標之採購標案有無共同特徵、涉及圍標案件的廠商、政府採購標案與招標機關有哪些組合及 LDA 主題分析，分析項目說明如次：

原始特徵值	原始資料型態	轉換後特徵值	轉換後資料型態
招標機關	字串	招標機關層級	類別(有序)
預算金額	數值	採購級距	類別(有序)
資本額	數值	廠商資本級距	類別(有序)
廠商地址	字串	廠商登記市縣	類別
機關地址	字串	機關所在市縣	類別

表 2. 特徵值轉換簡表

級距 標的	巨額採購	查核金額	公告金額	未達公告金額	小額採購
工程	2 億元	5,000 萬元	100 萬元	100 萬元以下	10 萬元以下
財務	1 億元	5,000 萬元			
勞務	2,000 萬元	1,000 萬元			

表 3. 政府採購級距認定標準表

- **分析一、違反採購法 87 條（圍標）的廠商有無共同特徵**：透過正規表達式抓取判決書中提及的圍標廠商名稱。使用商工行政資料獲取廠商資料，包括資本額、登記地區、組織類型等。進行特徵工程，並為了找出經常性一起出現之項目，本研究選擇 Apriori 適用於在資料集中找出其中的頻繁組合；再者因本研究之資料量較小，故 Apriori 演算法於資料量較大之資料集所造成的效率低下缺點並不會影響本研究。
- **分析二、發生圍標之採購標案有無共同特徵**：因本研究所使用之資料集為繁體中文，故使用支援繁體中文的 Jieba 進行斷詞並建立政府採購標案名稱字典，擷取判決書中的採購案名稱。整合決標資訊，包括招標機關、採購金額、標的分類等，進行特徵工程並運用 Apriori 演算法找出頻繁項集。
- **分析三、涉及圍標案件的廠商、政府採購標案與招標機關有哪些組合**：將分析一與分析二的資料合併，使用 Apriori 演算法找出涉及圍標案件的廠商、政府採

購標案與招標機關之頻繁組合，並將出現機率大於 0.50 者視為頻繁項目集。

- **分析四、LDA (Latent Dirichlet Allocation) 主題分群：**將經 Jieba 段詞後的涉及圍標案件之廠商、政府採購標案與招標機關資料合併，進行 LDA 主題分群，將其距離相近者分成五類，並找出其中出現次數最多的前 10 個項目。

透過此些分析可識別出相關的組合，進而提供政府審計人員在選案決策時的參考基礎，以辨識潛在高風險案件。

4 實驗結果

4.1 資料集

- **有罪判決書之過濾與統計結果：**根據刑事訴訟法 (314 之 1 條) 的要求，對有罪判決書進行過濾與統計。共有 447 件有罪判決書，透過正規表達式過濾並獲取了違反不同論罪條文的判決數量，如表 4 所示。

違反之法條	判決書件數	佔比
違反採購法 (87 條, 圍標)	435	97.32%
違反採購法 (92 條, 科廠商罰金)	296	66.22%
違反採購法 (88 條, 綁標)	12	2.68%

表 4. 有罪判決書之過濾統計結果

- **有罪判決書之招標機關擷取與統計結果：**利用 Jieba 工具結合政府機關名稱自定義字典，對 447 件有罪判決書中的政府機關名稱進行擷取與統計。共擷取出 853 個機關名稱，根據政府機關層級進行分類，如表 5 所示。
- **有罪判決書之廠商擷取與統計結果：**該部分使用 Regular Expression 擷取了 447 件有罪判決書中的廠商名稱，共獲得 1,483 個廠商名稱，分佈於 369 件判決書中。進一步透過商工行政開放平臺 API 取得公司、商號的登記資料，共獲得 1,054 個廠商的商工登記基本資料，如表 6 所示。

機關層級 機關類別	二級機關	三級機關	四級機關	五級機關
	中央政府	6	97	49
地方政府	153	251	34	0
公立學校	0	49	93	0
公營事業	0	61	54	6

表 5. 違反政府採購法之裁判書，擷取之政府機關資訊一覽表

資本額 級距	登記類型	公司	商業
	0 元至 5000 元		14
5000 元至 1 萬元		0	4
1 萬元至 10 萬元		5	7
10 萬元至 100 萬元		146	60
100 萬元至 1000 萬元		460	44
1000 萬元至 1 億元		270	0
1 億元至以上		39	0

表 6. 違反政府採購法之裁判書，擷取廠商資本額一覽表

- **有罪判決書之採購案名稱擷取與統計結果：**該部分利用 Jieba 結合採購案名稱自定義字典，對 447 件有罪判決書中的政府採購案名稱進行擷取與統計。共獲得 606 個採購案名，然而，由於有些案名相同導致串連了 1,670 個決標資料，因此進行了篩選，最終得到 586 個符合條件的採購案。根據採購標的類別分類，工程類採購最多。就採購金額規模而言，公告金額採購最多，未達公告金額之採購次之。招標方式方面，以公開招標為最多，限制性招標 (公開取得) 次之，詳如表 7 與表 8 所示。

採購金額規模	標的分類	勞務類	工程類	財務類
	未達公告金額		67	55
公告金額		38	190	108
查核金額		6	1	0
巨額採購		16	0	3

表 7. 違反政府採購法裁判書，擷取採購案之標的與採購金額規模一覽表

招標方式 \ 標的分類	勞務類	工程類	財務類
	公開招標	63	192
限制性招標 (公開取得)	54	46	97
限制性招標 (公開評選)	2	0	0
限制性招標 (未公開評選)	8	8	3

表 8. 違反政府採購法裁判書，擷取採購案之標的與招標方式一覽表

4.2 實驗分析結果

根據分析一、分析二以及分析三之結果，可以綜合歸納出以下有關違反採購法（87 條，圍標）的廠商特徵、發生圍標之採購標案特徵，以及涉及圍標案件的廠商、政府採購標案與招標機關組合的綜合觀察。

4.2.1 分析一、違反採購法 87 條（圍標）的廠商有無共同特徵：

- 公司種類：多為有限公司。
- 登記資本額：主要集中在 100 萬元至 1,000 萬元之間，也有一部分位於 1,000 萬元至 1 億元之間。
- 營業項目：主要涵蓋國際貿易業，少數涉及機械安裝業和室內裝潢業。

故研究分析結果顯示公司種類為有限公司且資本額為 100 萬元以上、營業項目主要國際貿易業者出現頻率較高。詳如表 9 所示。

頻繁項目集合	出現機率	頻繁項目集合	出現機率
['最低標']	0.87	['公開招標', '公告金額']	0.71
['公開招標']	0.78	['公告金額', '最低標']	0.65
['公開招標', '最低標']	0.72	['公開招標', '公告金額', '最低標']	0.65
['決標占底價 9 成']	0.72	['最低標', '決標占底價 9 成']	0.61
['公告金額']	0.71	['三級機關']	0.58

表 9. 違反採購法（87 條）廠商特徵頻繁項集

4.2.2 分析二、發生圍標之採購標案有無共同特徵：

- 決標方式：主要使用最低標方式進行決標。
- 招標方式：常以公開招標方式進行。
- 決標金額比例：決標金額通常占底價金額的 9 成以上。
- 招標機關：三級機關較常涉及。
- 採購標的：以工程類為主要採購標的。
- 投標廠商數：通常有三家合格廠商參與投標。

故研究分析結果顯示投標廠商數達三家、使用最低標或公開招標方式者進行決標、決標金額比例占底價金額的 9 成以上且涉及之招標機關為三級機關，並以工程類為主要採購者出現機率較高。詳如表 10 所示。

頻繁項目集合	出現機率	頻繁項目集合	出現機率
['最低標']	0.87	['公開招標', '公告金額']	0.71
['公開招標']	0.78	['公告金額', '最低標']	0.65
['公開招標', '最低標']	0.72	['公開招標', '公告金額', '最低標']	0.65
['決標占底價 9 成']	0.72	['最低標', '決標占底價 9 成']	0.61
['公告金額']	0.71	['三級機關']	0.58

表 10. 違反採購法（87 條）廠商特徵頻繁項集

4.2.3 分析三、涉及圍標案件的廠商、政府採購標案與招標機關有哪些組合：

研究分析結果顯示廠商種類多為有限公司，參與以最低金額決標的政府採購案或參與公開招標的政府採購案，且多以公開方式辦理 100 萬元以上（公告金額）的政府採購案，並以最低金額決標者出現機率較高。詳如表 11、表 12 與表 13 所示。

4.2.4 分析四、LDA (Latent Dirichlet Allocation) 主題分群

- 主題一的前十個常出現的詞彙涉及地方政府、學校和家具等相關主題，分析結

頻繁項目集合	出現頻率	頻繁項目集合	出現頻率
['有限公司', '最低標']	0.75	['100 萬元至 1000 萬元', '最低標']	0.66
['公開招標', '公告金額']	0.71	['公告金額', '最低標']	0.63
['公開招標', '最低標']	0.71	['有限公司', '公告金額']	0.61
['有限公司', '100 萬元至 1000 萬元']	0.71	['決標占底價 9 成', '最低標']	0.60
['有限公司', '公開招標']	0.68	['有限公司', '決標占底價 9 成']	0.60

表 11. 2 項發生圍標之採購案及廠商特徵頻繁項集之組合

頻繁項目集合	出現機率	頻繁項目集合	出現機率
['公開招標', '公告金額', '最低標']	0.63	['有限公司', '公告金額', '最低標']	0.55
['有限公司', '100 萬元至 1000 萬元', '最低標']	0.62	['100 萬元至 1000 萬元', '公開招標', '最低標']	0.55
['有限公司', '公開招標', '最低標']	0.62	['公開招標', '決標占底價 9 成', '公告金額']	0.52
['有限公司', '公開招標', '公告金額']	0.60	['100 萬元至 1000 萬元', '公開招標', '公告金額']	0.52
['有限公司', '100 萬元至 1000 萬元', '公開招標']	0.57	['公開招標', '決標占底價 9 成', '最低標']	0.51

表 12. 3 項發生圍標之採購案及廠商特徵頻繁項集之組合

果顯示其地方政府機構、學校設施或家具採購等距離較近。

- 主題二的前十個常出現的詞彙涉及土木工程、建築公司和地方政府機構等相關主題，分析結果顯示土木工程承包、建築業務和地方政府合作等距離較近。
- 主題三的前十個常出現的詞彙涉及教育、工程、政府項目和公司業務等相關主題，分析結果顯示公立學校改善項目、企業合作、政府委外項目和開發計劃等距離較近。

頻繁項目集合	出現機率	頻繁項目集合	出現機率
['有限公司', '公開招標', '公告金額', '最低標']	0.55	['最低標', '有限公司', '100 萬元至 1000 萬元', '公告金額', '公開招標']	0.45
['有限公司', '100 萬元至 1000 萬元', '公開招標', '最低標']	0.52	['有限公司', '100 萬元至 1000 萬元', '公告金額', '最低標']	0.45
['有限公司', '100 萬元至 1000 萬元', '公開招標', '公告金額']	0.49	['100 萬元至 1000 萬元', '有限公司', '三級機關', '最低標']	0.42
['100 萬元至 1000 萬元', '公開招標', '公告金額', '最低標']	0.48	['有限公司', '公開招標', '國際貿易業', '最低標']	0.42
['公開招標', '決標占底價 9 成', '公告金額', '最低標']	0.47	['公開招標', '有限公司', '決標占底價 9 成', '公告金額']	0.42

表 13. 4 項發生圍標之採購案及廠商特徵頻繁項集之組合

- 主題四的前十個常出現的詞彙涉及營造業、木材加工業、機關單位等相關主題，分析結果顯示營造公司、木材加工業務以及政府機關等距離較近。
- 主題五的前十個常出現的詞彙涉及到鋼鐵業、儀器製造、政府機構以及遠距居家照顧系統等相關主題，分析結果顯示適用於經營法令非禁止或限制的業務領域等相關領域距離較近。

故研究分析結果顯示地方政府機構可能與土木工程、學校設施或家具採購、鋼鐵製造以及營造業等產業類別距離較近；公立學校改善項目、企業合作、政府委外項目和開發計劃亦是出現次數較高之組合。詳細之分類結果詳如表 14 所示。

4.3 綜合討論

綜合以上分析，可看出違反採購法（87 條，圍標）的廠商在公司種類、登記資本額和營業項目方面有一定的共同特徵。發生圍標之採購案亦有明顯的決標方式、招標方式和招標機關之共同特徵。而涉及圍標案件時，一些特定的組合情況也常見於實際案例中。此些結果對於進一步的採購法規劃、執行和監管提供了一定之價值參考。然而，在研究過

主題一	主題二	主題三	主題四	主題五
三級機關	地方政府	公立學校	四級機關	得經營法令非禁止或限制之業務
彰化縣芬園鄉公所	南投縣南投市公所	台灣中油股份有限公司	三級機關	code
國立豐原高級商業職業學校	53482370	100 年第 1 階段改善無障礙環境工程	慕昀營造有限公司	鋼鐵軋延及擠型業
90	土木包業	雲林縣元長鄉公所	2012	儀器
地方政府	三級機關	臺北縣板橋市公所	100	中央政府
家具	臺北縣板橋市公所	103 年度收費管理及設備維護勞務委外案	營造業除外	65
宜蘭縣南澳鄉公所	弘祥土木包工業	澎湖縣立文光國民中學	800000	b008
高雄縣甲仙鄉公所	名慶營造有限公司	茂富有限公司	木材製品加工製造及買賣業務	高雄市甲仙區公所
研究	國軍退役役官兵輔導委員會	開發	103	104 年遠距居家照顧系統
嘉義縣政府	平治營造股份有限公司	8196721311475	功和營造有限公司	24441226

表 14. LDA 主題分群結果

程中亦有些困難和限制。如資料清理過程，特別是裁判書附表的清理工作，耗費了大量的人力和時間。廠商名稱的擷取和公司登記資料的匹配存在同名和模糊匹配的問題可能影響了分析的精確性。此外，頻繁項集探勘結果的通用性特徵可能減少了對其他更特殊特徵的檢測。以及 LDA 分群結果因採用非監督式學習，無法如人為判別般精準分類。

5 結論

本研究運用 Jieba 斷詞技術、頻繁樣式探勘技術與 LDA 主題分群，對違反政府採購法案件進行了特徵探勘。使用 Jieba 斷詞技術對判決書進行段詞；頻繁樣式探勘技術找初出現機率大於 0.50 之頻繁項集；LDA 主題分群出距離相近之五大主題，並找出其中出現次數最高之幾組項集以輔助審計人員選擇具有高風險形態的政府採購案件。儘管在研究過程中含有部分限制，然而本研究仍為政府審計工作提供了有價值的參考資料。未來的研究可考慮採用多元方法，如社群網絡分析，以揭示廠商間、廠商與招標機關間，以及廠商與採購案之間的關係，並結合其他資料源來提高分析的準確性和可靠性。進一步的特徵工程和機器學習模型的應用有助於提升分析的精確性和預測能力。最後，擴大研究範圍和時段將有助於獲得更全面且代表性的結果。

總體而言，本研究為政府採購案件中的圍標現象提供了洞察和分析，並為未來的研究提供了指引。進一步的改進和擴展將有助於更好地協助審計人員選擇和稽查高風險形態的政府採購案件，並為審計工作提供更有效的工具和參考。

Acknowledgments

The research work described in this paper was supported by the grant from the National Science and Technology Council (NSTC), Taiwan, ROC, under Grant No. MOST 111-2221-E-031-004-MY3.

References

馮世墩(2018)。公共工程採購刑事責任之研究—以圍標與綁標行為為核心。東吳大學法學院法律學系碩士在職專班法律專業組碩士論文。台北市。取自 <https://hdl.handle.net/11296/rh8eqp>

- 廖鼎銘(2004)。觸犯多款法條之賭博與竊盜案件的法院文書的分類與分析。國立政治大學資訊科學學系碩士論文。台北市。取自 <https://hdl.handle.net/11296/7uy772>
- 何君豪(2007)。階層式分群法在民事裁判要旨分群上之應用。國立政治大學資訊科學學系碩士論文。台北市。取自 <https://hdl.handle.net/11296/6kp6g2>
- 黃玉婷(2012)。以文字探勘技術產製求/量刑因子之研究—以我國智慧財產權法律為中心探討。東吳大學法律學系碩士在職專班科技法律組碩士論文。台北市。取自 <https://hdl.handle.net/11296/3spke3>
- 陳政瑜(2015)。基於文字探勘技術探討司法裁判書之撰寫一致性:以刑事訴訟停止羈押聲請裁定書為例。國立臺灣大學管理學院資訊管理學研究所碩士論文。台北市。取自 <http://dx.doi.org/10.6342/NTU.2015.02717>
- 林筱鈺(2013)。文字探勘在判決書上之應用-以著作權法民事賠償為中心。國立臺灣大學法律學院科際整合法律學研究所碩士論文。台北市。取自 <http://dx.doi.org/10.6342/NTU.2013.00586>
- Chee, C. H., Jaafar, J., Aziz, I. A., Hasan, M. H., & Yeoh, W. (2019). Algorithms for frequent itemset mining: a literature review. *Artificial Intelligence Review*, 52(4), 2603-2621.
- Li, Y., Krishnamurthy, R., Raghavan, S., Vaithyanathan, S., & Jagadish, H. V. (2008, October). Regular expression learning for information extraction. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing* (pp. 21-30).
- 黃詩淳、邵軒磊等人(2018)。酌定子女親權之重要因素:以決策樹方法分析相關裁判。臺大法學論叢, 第 47 卷第 1 期。取自 <https://www-airitilibrary-com.eu1.proxy.openathens.net/Publication/alDetailedMesh?DocID=a0000446-201803-201804100003-201804100003-299-344&PublishTypeID=P001>
- Alexander V. Mamishev and Sean D. Williams. 2010. *Technical Writing for Teams: The STREAM Tools Handbook*. Wiley-IEEE Press, Hoboken, NJ.
- Mohammad Sadeqh Rasooli and Joel R. Tetreault. 2015. *Yara parser: A fast and accurate dependency parser*. *Computing Research Repository*, arXiv:1503.06733. Version 2.
- Hui-Hsin Tseng, Chao-Lin Liu, Zhao-Ming Gao, and Keh-Jiann Chen. 2002. 以構詞律與相似法為本的中文動詞自動分類研究 (a hybrid approach for automatic classification of Chinese unknown verbs) [in Chinese]. In *International Journal of Computational Linguistics & Chinese Language Processing, Volume 7, Number 1, February 2002: Special Issue on HowNet and Its Applications*, pages 1–28.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding. *Poetics*, 41(6), 570-606.
- Calheiros, A. C., Moro, S., & Rita, P. (2017). Sentiment classification of consumer-generated online reviews using topic modeling. *Journal of Hospitality Marketing & Management*, 26(7), 675-693.