

## 台灣立法委員政見資料庫之建置\*

廖達琪\*\*、林福仁\*\*\*、黃郁慈\*\*\*\*、劉子昱\*\*\*\*\*、李承訓\*\*\*\*\*

### 《本文摘要》

台灣的國會選舉中，一向規定個別立法委員候選人需提出政見，並刊載在選舉公報上。雖然目前坊間主流意見多認為政見，尤其登之於公報上的政見，一般選民根本不看，與候選人當選與否亦應是關係不大，所以相關的研究並不多。但本研究認為公報上的政見，既是候選人正式的選前承諾，不應積非成是，在研究上就理所當然予以忽視；反而應藉由探討這些承諾與選後國會中表現的聯結情形，來強化選民對代表的「課責性」；也希望因而能促進各方對公報上「政見」的重視，形成民主政治的良性循環。

爲了達到上述目標，本文針對我國第六、第七屆立委選舉政見，建構「第六及第七屆立委政見資料庫」，希望藉此資料庫的建置，未來探討如立委選前政見與選後立法表現等相關研究議題。

本文主要說明此資料庫的建構過程，包括所用的人工登錄方式和資料採礦的資訊技術，特別是新進發展的多重標籤分類方式；本文並企圖透過這兩種方式的比對，未來能引進資訊技術，對每一次選舉的公報政見，能做相當程度的自動化分類處理。

關鍵詞：立法院、立法委員、政見資料庫、資料探勘、多重標籤分類

\* 本論文感謝國科會專題研究計畫「國會議員競選承諾與選後立法表現」(NSC98-2410-H-110-015-MY3)補助支持，更感謝審查人的寶貴意見；一切文責作者自負。

\*\* 國立中山大學政治學研究所教授。E-mail: dcliao@mail.nsysu.edu.tw。

\*\*\* 國立清華大學服務科學研究所教授。E-mail: frlin@mx.nthu.edu.tw。

\*\*\*\* 國立清華大學科技管理研究所碩士。E-mail: whoissofa@gmail.com。

\*\*\*\*\* 國立中山大學政治學研究所博士生。E-mail: tsuyuliu@yahoo.com.tw。

\*\*\*\*\* 國立中山大學政治學研究所博士生。E-mail: m946020014@student.nsysu.edu.tw。

## 壹、前言

自由、公開、定期的選舉，是民主國家的基本運作方式。政黨及政治菁英經由選舉的過程，讓人民決定誰可以獲取分配公共利益的權力，並且在法定任期結束後，交由公民們以選票決定政治菁英的去留。參與公職選舉的政黨及候選人們，無不希望能獲得選民大多數選票的支持，以便順利取得執政的權力。

爲了能爭取最大多數的選票，政黨及候選人必須向選民宣告其政治理念，端上各式各樣的政策牛肉，以便選民能夠在競爭激烈的選戰中，挑選到理念最接近的政黨或候選人。在大多數的民主國家中，候選人的競選政見，在選舉前必需公開接受選民以及競選對手的批評與檢驗，並在當選後於任期內盡可能的將其競選承諾予以兌現，否則可能會在下一次的選舉中遇到連任的麻煩。此處所謂的「競選承諾」(campaign promises)，泛指候選人在競選過程中，向選民公開保證 (pledge) 會完成或履行的任何事項，目前大約有下列三種方式呈現：

- 一、政黨政綱 (party platform)：民主國家的政黨，通常都會擬定一份「政黨政綱」，內容可能包含：該政黨的意識型態、成立政黨目的、政黨在不同政策領域的理念或是該政黨的政治行動準則。以美國爲例，政黨會在一段時間內修改或更新其部分的政黨政綱，修正該政黨在某些領域的政策方向。而該政黨角逐選舉席次者，如總統或國會議員候選人則會公開宣示支持其政黨政綱。現今歐洲的民主國家，亦大都採用政黨政綱方式，讓選民選前有所參考選擇。
- 二、候選人政見承諾：除了政黨政綱外，候選人也會透過新聞媒體、廣告、演講、宣傳、座談會，以及各式集會，向選民宣告其執政理念、政策方向、預算分配等政見承諾，此類的宣示政見方式，泛見於大多數的民主國家選舉活動。
- 三、候選人政見公報：少數國家會正式規定，候選人在登記參選時，必須同時提出其競選政見，並登錄於正式的競選公報之中，提供選民做爲投票時的參考；我國及日本即採此方式。以我國的〈公職人員選舉罷免法〉爲例，在第四十七條中明文規定選舉公報應包含：「區域、原住民立法委員及地方公職人員選舉，各候選人之號次、相片、姓名、出生年月日、性別、出生地、推薦之政黨、學歷、經歷及政見……政見內容以六百字爲限。」

以上這三種型式的競選承諾中，第一種的政黨政綱被研究探討的最多，並在歐洲建立了政黨政綱的資料庫 (comparative manifestos project, CMP)，藉以做跨國比較政綱及政策的關聯性等議題；第二及第三種的相關研究則相對較少。本文認爲：以規範民主理論的視

角而言，競選前的承諾，無論以何種形式出現，都可視為是與選民的約定，有被選前檢視及選後追蹤是否兌現的必要。所以就民主實踐的研究探討而言，三種競選承諾的重要性無分軒輊；但就實務上，相關政見或承諾可蒐集的系統及完整性而言，第一及第三種遠優於第二種；再就與選民約定的正式及公開性而言，第一及第三種亦是遠優於第二種。只是相當遺憾，現有文獻對第三種政見的整理蒐集及分析，是相當缺乏的，尤其是這種公報政見與當選後的施政或立法表現之關聯，在現有文獻中的探討，亦甚為稀薄，這也構成本文撰寫的主要動機。

雖然目前台灣坊間的看法，較傾向認為公報上的政見不重要，一方面選民很少參考做為投票依據；另一方面，這些政見內容五花八門，恐怕形式意義大於實質，似不值得探討其內涵或與政策的關聯性等等。本研究了解這樣質疑的存在，但有不同之見解及實務資料之依據，因為就規範民主理論意涵而言，如前所述，官方正式登錄的政見，應有契約之性質，選民重視與否，不是重點。研究者追蹤了解，協助選民課責 (accountable)，具有理論及實務上的正當性；再就政見的內容而言，經過實際的審視台灣歷屆選舉公報，牛肉成分實不少，值得分析探討。

本文乃做為探討公報政見的第一步，主要聚焦於如何建立資料庫，並以台灣第六及第七屆立法委員候選人為對象，以他們登錄在中央選舉委員會所出刊的選舉公報上之政見為內容，再結合人工登錄方法及資料探勘技術，希望逐步發展出以資訊技術來做自動化登錄政見的一套方法，以為將來繼續分類登錄公報上選舉政見之用。

本文計分五部分：壹為此前言，貳為政見相關研究之回顧及前瞻，參為政見資料庫之建構——人工登錄過程說明，肆為資料探勘技術應用——多標籤文件分類法 (multi-label classification approach)，伍為結論。

## 貳、政見相關研究之回顧及前瞻

如前所述，有關政黨政綱的研究是文獻中的較大宗，所以有關政見的文獻回顧，即以此為主要內容，以下分兩部分陳述：

### 一、政黨政綱位置與民意及其測量方法

當代幾個以理性選擇研究途徑為基礎的理論，如趨近模型 (proximity spatial model)、方向模型 (directional spatial model) 與顯著性理論 (saliency theory)，都堅信選民會投票給在政策空間上位置較接近自己立場的政黨 (Budge 2001; Downs 1957; 胡佛 1986; 徐火炎 1991; 1993; 劉從葦 2006)。此類理論認為，政黨或候選人必然會擬定較符合其支持

者，或其所屬地區選民需求的政見，以爭取政策空間位置較接近的選民投票支持。此外，Alan D. Monroe 在 1983 年出版的 *American Party Platforms and Public Opinion* 文獻中，則是運用民意調查單位所做出的民調，來檢驗黨綱的內容是否會隨著民意而有所變動，其研究顯示黨綱與民意有著相當高的關聯性 (Monroe 1983)。

測量政黨位置的方式又可分為兩類：政見內容分析法與專家調查。前者如 1979 年成立的政見研究小組 (Manifesto Research Group, MRG)，以及 Roberson(1976)、Budge 與 Farlie(1983) 等研究團隊為主，將政黨政見細分為不同的類別。MRG 小組是經過各國研究政黨的學者討論過後，將政黨政見區分為七大政策領域、54 個政策類別的編碼架構。Roberson 以及 Budge 與 Farlie 則是以問卷的形式，將政見及問卷寄給相關領域的專家學者，請他們為各政黨定位其空間位置。

由於各國的政治制度有異，政黨及候選人提出政見承諾的主要方式也有所不同，例如歐洲國家以政黨政見為主，而台灣及日本則是候選人自行提出個人政見為主。因此這類研究又可區分為，以政黨政見定位政黨位置，以及以政黨所提名候選人的政見測量政黨位置兩種。前者例如上述所提及的 MRG 小組，後者則如我國學者劉從葦，其根據 MRG 的分類架構，因應我國的特殊政治制度及情勢，另增加了 14 個其他的政策類別，從而分析我國各政黨及政黨的各派系在政策位置上的不同，並發現政黨政見與以選舉公報為文本的內容分析所呈現的政黨相對位置大致相同 (劉從葦 2006)。

## 二、對政見履行程度的分析

此類型的研究認為，基於民主政治所重視的課責性與對民意的回應性等核心價值，經由民主程序產生的當選者必然有義務盡可能的履行其競選承諾。事實上，若是當選者無法在任期內完成其大部分的競選承諾，也有可能面臨連任的困難，因為挑戰在職者的對手總是在尋找現任者的缺失 (Arnold 1990; 2004; Sulkin 2012)。

對於競選承諾的研究顯示，當選者也確實在其任期間，盡可能地履行其競選承諾。例如 Pomper(1968; 1980) 整理了 1944 到 1964 年間美國兩黨的黨綱，並將黨綱的內容予以分類，主要分成：口惠與事實的 (Rhetoric and facts)、評估政黨施政紀錄及表現 (Evaluations of the parties' records and past performances)、未來政策的論述 (Statements of future policies) 三大類型。<sup>1</sup> 以上述分類檢視總統當選人，履行政黨黨綱所提出的競選承諾之比例，

<sup>1</sup> 在評估政黨施政紀錄的類型中，分為：普遍認同 (General approval)、普遍批評 (General criticism)、政策同意 (Policy approval)、政策批評 (Policy criticism) 四個項目；而在未來政策的論述類型中，分為：冠冕堂皇型承諾 (Rhetorical pledges)、普遍型承諾 (General pledges)、延續型承諾 (Pledges of continuity)、表達關注型承諾 (Expressions of goals and concerns)、實踐型承諾 (Pledges of action)、細節型承諾 (Detailed pledges)。

Pomper 估計這段期間總統當選人所完成的政見比率約在 2/3 至 3/4 之間。

相對於以上這些政黨政綱的分析及履行的探討，我國在政黨及候選人政見履行的研究上則顯得較為缺乏；雖然，近年來針對選舉公報上的政見，也有一些專門議題性的分析，如檢視立委選舉的醫療保健政見之專業性及合法性（魏中仁與楊志良 1997），或探究立委的體育政見類型和當選的關係（林季福 2003）等；但目前僅有黃國敏（2009）曾針對縣市長層級的政見履行情形做個案研究之實證分析，以及林東穎（2010）就立法委員在國會中的立法表現，探討立委的政見實踐比例。

本文認為，政見在民主國家的政府運作、公職選舉與利益團體及一般公民之間，扮演了重要的接合角色。特別是，我國選罷法第 47 條明訂，選舉公報需載明候選人政見內容至少六百字。因此選舉公報上的政見除了應視為候選人必須履行的正式承諾外，學界也應著手建立可有效檢視當選者如何履行政見承諾的方法及準則；更何況，公報政見的研究實可協助我們釐清一些困惑及帶動一些重要的研究議題，試舉例如下：

（一）公報政見有哪些類型？台灣特有的選舉公報上登錄的「政見」到底出現了那些承諾？它的大體結構類型又為何？雖然我國的選罷法對候選人的政見內容有字數的限定（600 字），但候選人在選舉公報上所載的政見內容卻相當繁複，包括：國家定位、國際外交、政府組織、中央預算、司法改革，乃至地方建設、對單一政治人物的支持、候選人個人特質的介紹等等；如此洋洋灑灑五花八門的政見，該如何分門別類整理？又與實際的施政或問政有何關聯？

（二）候選人當選與否，和公報政見的性質、地域方向、議題類別等有無關聯性？候選人提出政見的目的，多是為了增加勝選的機率，但實際上的效益如何？目前文獻有不同看法，有的認為有幫助（廖益興 2005）；有的不同意，認為是聊備一格（陳永峰 2005）；有的則認為政見對不同層級選舉的效益是有差別的：縣長層級的選舉，政見的重要性似優於立委層級（王鼎銘 2003）。這些不同的研究結論，均非得來自整體大量的候選人公報政見分析與實際選舉結果的聯結，所以公報政見資料庫的建立，多少亦有助於對此議題進一步的釐清。

（三）順利當選的候選人，會否實現其競選承諾？如何履行？如前所述，國外的文獻顯示順利當選的候選人會盡量履行其競選承諾，使其政黨及個人在未來選舉中能獲得較多的優勢（Elling 1979）。相對於國外，我國的候選人必須在選舉公報中，白紙黑字註明其政治理念及競選承諾，這樣的制度會否影響到其當選後的作為？以立委為例，當選者是否會因此在立法院的議事過程中，較為關注和其政見有關的議題？又是否會努力推動和政見有關的議案？

本文認為，公報政見資料庫的建立，可至少帶動探討上述的三項議題，所以先選定台

灣第六 (2005-2008) 及第七屆 (2008-2012) 所有立委候選人在選舉公報上的競選政見，做為登錄對象。至於縣長層級的政見，因已有研究者涉入，本研究即不處理。而總統層級的政見，因參與人少，舉辦屆數又不多（迄今四屆而已），還不到處理的時候。在立委這一層級，則尚未有研究全面處理過。且不論從民主理論或實踐的面向上看，了解立委在公報上的正式承諾，均有重大意義。而特別以第六及第七屆為對象，則主要是台灣的選舉制度在第七屆有重大變革，從原來的複數選區單記不可讓渡投票制 (MMD and SNTV)，改為單一選區兩票制 (SMD and PR)。這樣的變革，是否會影響政見提供的地域方向及內涵等，又是一個實務及理論均有意義的研究議題。

根據上述之研究背景，本文針對我國第六屆及第七屆的立法委員選舉，所有候選人在選舉公報所刊登的競選政見，做一全面性的政見資料庫建構，並由廖達琪教授所率之團隊針對立委候選人之政見文本，做人工的政見分類，再將團隊所設定之政見分類規則、方法及編碼完成之資料庫，交由清華大學林福仁教授團隊以「文本探勘法」(text-mining) 為基礎的研究軟體學習人工編碼之思維；並以第六屆的資料為學習樣本，試著分類第七屆之政見，再與人工所登錄之結果做比對，期望能夠以資訊科技與政治學結合的方式，為未來的立委選舉政見做自動分類，以及進一步檢視立委政見與其在立法院表現的關聯。本文接續將詳述此一政見資料庫的建構過程。

## 參、政見資料庫之建構——人工登錄過程說明

立委政見資料庫的人工建構過程，可分為三個部分說明，分別為：一、政見分類及定義；二、巢狀編碼；及三、信度檢核；以下分別說明：

### 一、政見分類及定義

在建立立委政見資料庫之前，必須針對政見進行分類。以往的研究中，針對政黨的政綱、候選人的政見做分類或分析的文獻不多。除了前述的 Pomper 將政黨政見分為三大類外，另 Terri Susan Fine 亦提及有關政見 (platforms) 內容分析的重要性 (Fine 1994)，但未曾實做；MRG 則將政黨政見區分為七大政策領域、54 個政策類別的編碼架構。

上述的政策領域與政策類別分類，較適用於測量政黨或派系政見的相對位置，若是要分析我國個別候選人的競選政見，分析候選人政見和其當選的關聯性，以及將建構完成的政見資料庫用於檢視當選者日後在立法院的表現，以及是否兌現競選政見等等，則必須重新考量訂做一個較詳細的政見分類標準。

以我國為例，許多候選人都會在政見中宣示個人的特質，以及強調會為地方爭取相當

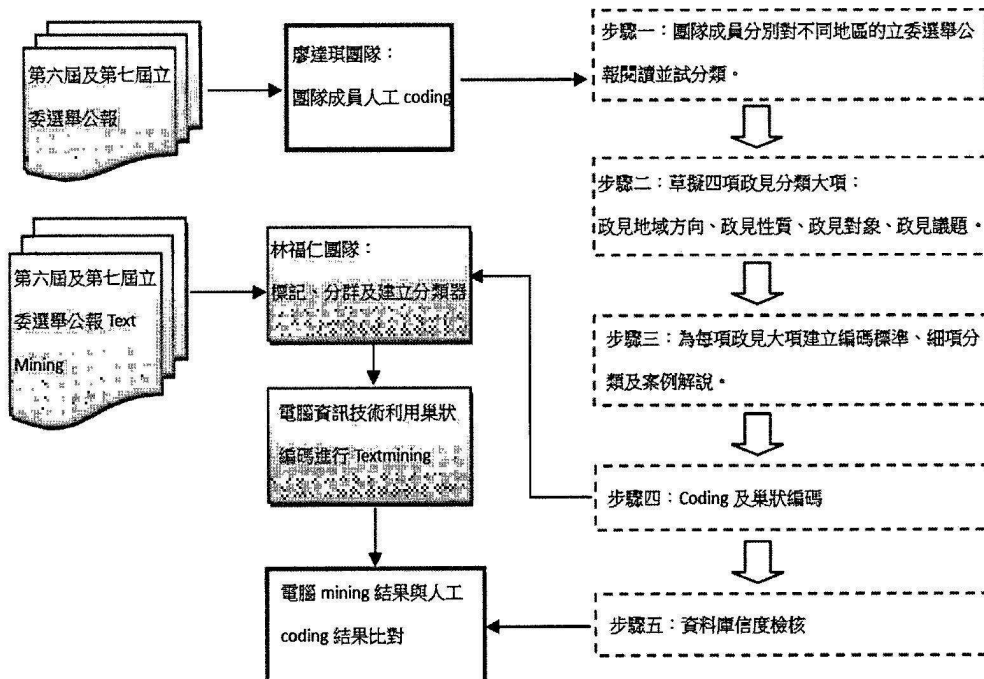
數量的地方建設；而即使是提及同樣的政策性政見，有些候選人能夠具體提出完整的政策規劃及配套；但有些候選人則提不出較具體的內容，如：「友善高雄、幸福台灣」、「理性制衡、全力改革」、「壯大台灣，中華民國再出發」等較為口號性的政見，就很難劃分至 MRG 小組或劉從葦所提的政策分類當中。

此外，MRG 小組的政策分類，也並未區分政見的地域方向、性質、對象及所用詞彙的意涵。在我國的特殊政治環境下，候選人通常會提出許多不同「地域方向」的政見，例如國際方向、兩岸方向、或是地方方向、跨區方向；也會針對不同的「對象」提出不同的政見，例如農漁民居多的鄉村地區，候選人就會針對農漁民的福利提出不同的政見規劃；原住民候選人則理所當然的提出較多以原住民為受惠對象的政見。若是單以 MRG 小組的政策分類，將沒有辦法分析到選民是否會有特殊的政策偏好，也無法比較同選區甚至是同黨候選人間的政見差異性。

有鑑於台灣的選舉公報制度、政治文化的不同，以及未來對立委於立法院中立法表現追蹤之考量，研究團隊重新設計一適用於台灣的政見分類方式。建構政見分類的方式及過程如下圖 1：

資料來源：本研究。

圖 1 政見編碼流程圖



研究團隊形成後（總計 10 位政研所研究生），在建立政見資料庫前，先將第六與第七屆的立委政見公報交由研究團隊的成員閱讀，並由主持人及成員分別試擬分類類別。經過

五個月的試分類以及細節討論，研究團隊草擬了四個政見分類大項，分別是：政見地域方向、政見性質、政見對象及政見議題類別。此四大項領域，是針對政見本身的屬性所進行的分類，其中，政見對象及政見議題分類，在每次的開會及試分類的過程中，都會針對無法歸類的案例，不斷地進行新增或修正，並將對象項目控制在 30 項，同時政見議題類別則控制在 14 項。為了讓每個團隊成員都有共同的編碼標準，各個不同的議題分項，都有具體的說明以及分類規則。

每個政見分類領域下，則另有更詳盡的政見次分類，由研究團隊成員共同討論決定之。以下分別說明研究團隊對四大政見領域與每個領域下的定義、次分類及分類原則：

### （一）政見地域方向

所謂「政見地域方向」，指候選人所提的政見，其內容涉及的地域範圍為何，並依地域範圍分為：全國性、單一地方、跨區性、兩岸以及國際。

民主國家的國會議員分別就全國以及所屬選區（單一地方）發表政見，應屬正常；而我國許多立法委員也會針對國際外交事務（如參與聯合國、世界衛生組織），以及兩岸交流（如三通、軍事競爭）等議題表達己身立場。因此研究團隊將立委的政見，在「政見地域方向」的類別上，初步區分為：全國、單一地方、兩岸與國際。這樣的分類方式，馬上遇到兩項問題：

1. 許多政見中，候選人雖提出了地方建設、地方規劃的政見，但該政見所及地域並不侷限於候選人所屬的選區或是單一縣市。例如：整治基隆河、修繕中橫公路等，這類的政見所涉及的地域都是跨縣市的。為因應此問題，團隊乃另增加了「跨區」這一項政見類別。
2. 因為此一類別是瞭解候選人的主要政見地域方向，研究團隊規定每一項候選人的政見中，只能歸類到一個政見地域方向之中，事實上，絕大部分的候選人的每一項政見，也只會涉及一個地域方向。然而，研究團隊成員在多次的模擬分類過程中，發現有少數的候選人政見會有「多重政見地域方向」的問題。例如，第七屆台中縣候選人簡肇棟曾提出「以台灣名義加入聯合國及世界衛生組織，發展兩岸和平方案。」在此政見中，同時出現了「國際」及「兩岸」兩項政見地域方向。研究團隊經過討論後，做出一項共識的規定：「當同一項政見涉及不同地域方向時，將此政見歸類在範圍最大的政見地域方向」。以簡肇棟的政見為例，即是將此一政見的地域方向分類在「國際」之中。表 1 為「政見地域方向」類別的說明及定義表。

表 1 政見領域 (一)：政見地域方向說明

政見地域方向	
1. 每一個政見只會有一項「政見地域方向」。	
2. 同一項政見涉及不同地域方向時，將此政見歸類在範圍最大的地域方向。	
政見地域方向	定義與規則
(1) 全國性	政見內容屬全國普遍適用者，如：修憲、募兵等。
(2) 單一地方	政見內容僅適用於單一地方者，如：建設信義區、南港鐵路地下化等。
(3) 跨區性	政見內容橫跨不同選區（縣市）者，如：基隆河之整治、中橫公路之修繕、中部捷運系統延長計畫、爭取國道四號台中環線豐原大坑段。
(4) 兩岸	政見內容特別提及「台灣與大陸之各種形式的交流」者，如：反對兩岸軍備競賽。
(5) 國際	指政見內容為參與國際組織或其他與外國交流之事務，如：參加國際組織：聯合國、WHO、WHA、海洋台灣等。

資料來源：本研究。

## (二) 政見性質

「政見性質」指該政見的屬性為何，並細分為：非政治性口號、政治口號、政策意涵及個人特質。此分類的目的是，區分候選人是否有提出「具體且有規劃」的政見議題。

每一位候選人提出政見時，通常會在政見中宣告自己的政治立場、為地方爭取建設，以及彰顯自己和其他候選人有何不同的特質。若是提出的政見有清楚的目標及實踐手段的論述，或是政見中有指出制定或修正某個法律案、確實提及某一建設的進行方式，則研究團隊會將其歸類在「政策意涵」的分類當中。

若是該政見雖然提出了某種建設上的規劃或政策上的承諾，卻沒有提出相應的施行手段或具體目標，則這樣的政見不具上述「政策意涵」定義。此類型的政見，依是否涉及政治議題分為「非政治性口號」及「政治性口號」。非政治性口號，特別指涉關於民生或經濟性議題的政見，只是沒有提出比較具體實踐的方法，例如：「揮別苦日子、再創好生活」、「貼近人民心聲、效率責任自許」。而「政治性口號」，則指難以在公共政策層次上釐清，必須進行政治動員的政見（含：涉及國家領土主權的認定），例如：反貪腐、修憲、討黨產、進入聯合國、台灣國、去中國、支持軍購保衛台灣等屬政治性口號。

最後，許多的候選人，特別是尋求連任的現任立法委員，通常都會強調個人的特質，例如個人對基層的服務是否紮實、曾經為地方爭取了什麼建設、在國會殿堂中的表現如何等。這類屬於描繪個人形象或個人政治績效的政見，研究團隊將其認定為「個人特質」式的政見，其定義為：強調個人的服務精神、專業、年輕、或資深，或含個人過往政績的宣揚。由於候選人在提及個人特質時，這樣的政見通常不具有「地域方向性」，也不特別針

對某個族群或「對象」強調其服務品質，因此研究團隊在此項政見類別中特別訂定一項規則：當立委所提出的政見是「個人特質」的宣揚時，無需填寫「地域方向」及「政見對象」。表 2 為「政見領域（二）」：「政見性質」類別的說明及定義表。

表 2 政見領域（二）：政見性質說明

政見性質	
政見性質	定義與規則
1. 政見有「政策意涵」時，進一步將該政見區分為「對象」及「議題類別」。	
2. 非政治性口號及政治口號，不列入議題類別，但可以有對象。	
(1) 非政治性口號	民生或經濟議題，但是沒有比較具體實踐的方法或目標，或有說法沒做法。例「揮別苦日子、再創好生活」、「貼近人民心聲、效率責任自許」。
(2) 政治口號	難以在公共政策層次上釐清，必須進行政治動員的政見（含：涉及國家領土主權的認定）。例：黨產、貪腐、修憲、討黨產、進入聯合國、台灣國、去中國、支持軍購保衛台灣、反軍購、台灣主體、國家（領土、土地）資源共管入憲。
(3) 政策意涵	政見有比較清楚的政策目標及實踐手段的論述，或是政見中有指定制定或修正某個法律案、確實提及某一建設的進行方式。例如：修正黨產條例、廢除徵兵制、政府編列預算為婦女免費施打子宮頸癌疫苗、建構高雄都會區快速道路系統等。
(4) 個人特質	強調個人的服務精神、專業、年輕、資深，含個人過往政績的宣揚。例如：「精簡兵力：為使國防資源運用最佳化，李文忠完成「台灣兵力規模研究報告」促使國防部李傑部長對外宣布民國九十八年義務役役期將縮短為十個月。」此段政見主要是宣揚個人的過往政績，因此列為個人特質。再如，某立委政見為：「第一、不收紅包。第二、不拿回扣。第三不包工程。」 規則：當立委所提出的政見是「個人特質」的宣揚時，無需填寫「地域方向」及「對象」。

資料來源：本研究。

### （三）政見對象

不論是哪一層級的選舉，候選人都會有針對不同的群體的選民提出適應的政見。因此，研究團隊也建立了一項「政見對象」的分類，定義為：指政見中所明確提及的政見受惠對象，並依候選人政見所使用的對象詞彙分為：全民、軍公教、台商、老人、婦女、原住民、族群（含閩南、外省、客家、若政見中特別提出「眷村」則擺入「族群」）、外籍人士（含外籍配偶）、學生、中壯年、青少年（青年及少年）、兒童（含嬰幼兒）、榮民（含榮譽）、勞工（農漁林牧礦計藝）、工商企業、公益團體（含社福團體）、專業人士、特殊技能人士、少數（含性工作者、煙毒犯、更生人、卡奴、腳踏車騎士，及所用詞彙含「弱

勢」的字眼)、僑民、殘障(含身心障礙)、失業、低收入戶(貧窮家庭、債務家庭)、中間選民、投資者、父母家長親子、單親家庭、選任公職人員、農漁民以及網民等 30 項。

當然，立委候選人通常所提及的對象是非常的包羅萬象，並不會僅止於上述的 30 項。但上述的政見對象項目，都是為研究團隊成員在針對第六屆及第七屆立委候選人政見的多次模擬分類過程中，選出被立委候選人提及頻率較高，或能反映時代變遷特質的項目。前者如殘障(身心障礙)福利或老人福利，是許多候選人都會在政見中提及的；有些以農漁業為主的縣市，候選人更是會提出許多保障農漁民福利的政策性政見。後者如腳踏車騎士、網民，這兩個對象雖然相對於農漁民、老人或學生等，較是小眾群體，但由於近年來腳踏車運動的盛行，以及網路犯罪的比率愈來愈高，在第六屆及第七屆的立委候選人政見中，被當成訴求對象是相當具有意義的，因此研究團隊將這兩項分類納入政見對象當中登錄。

#### (四) 議題類別

議題類別是針對立委政見所提到的，具有「政策意涵」的政見，依該政策的「主管行政機關」及在立法院中的「責成委員會」所分類。

首先，由於立委候選人所提出的具有政策內涵的政見，極大部分都涉及「內政」。因此，在「內政」此一議題類別之中，我們進一步的區分出：地方、族群、警政治安、區域整合、政黨政治以及非上述範圍的「其他」。其中「地方」又再細分為「服務」、「藝文休憩觀光」、「基礎建設」及「地方建設」。內政之外，我們則進一步依**中央部會職掌及立法院各常設委員會功能**之規範分類，另增加「其他」這一議題分類。整體議題類別包括：憲政議題、國會改革、外交國防、經濟、財政、教育文化、交通、法制司法、福利、消費者保護、醫療衛生、環境及其他。關於以上分類的定義與案例請參考下表 3。

以立法院各常設委員會的功能職掌為分類依據有下列兩個原因：其一是，未來研究團隊要進一步分析立法委員在國會殿堂中，其問政及法案的提出是否有履行其競選承諾。而每位立法委員皆有其專業及所處的常設委員會，因此依立法院各常設委員會功能職掌建立分類標準，有助於研究團隊在未來能以同一的標準分析立法委員在國會的行為。其二，大多數的立委政見有賴中央行政單位執行，但中央部會的機關既多且雜，許多業務也會交互重疊，經常有跨部會處理的情形，若依中央行政單位功能，將立委政見分類，會造成分類過於細碎且分析時仍要做整合歸併處理。而換個角度看，每一中央部會都需要到立法院面對立委的問政，每一部會也都會有相對應的立院常設委員會，相較於中央行政單位，立院常設委員會的數目及功能就單純許多。<sup>2</sup> 因此，當我們在判斷一個政見的時候，先找出

<sup>2</sup> 立院常設委員會在第六屆為 12 個，第七屆則縮減為 8 個，數目遠較中央行政機關為少。而每一常設委員會，在立法院中都有清楚的功能職掌。

可能的主管中央部會，若是有多個部會都有可能處理到這項政見時，再進一步分析其在立院會分至哪一個常設委員會中，即可較清楚的判斷此一政見的分類為何。

在建立好議題分類的類別後，研究團隊進一步討論「議題類別」此一政見領域中的政見分類規則。首先，在「內政」此一議題類別中，因為內政的政策政見極多，所以有了許多進一步的細分。但在「基礎建設」以及「地方建設」這兩個項目中，大部分有關這兩項建設政策的政見，同時可能涉及其他政見議題分類，但「表面上」並無明顯標記。例如，「爭取在桃園縣建設巨蛋球場」這樣的一個地方建設的政見，會同時直接涉及體育事務，而體育事務又屬教育文化委員會的職掌，因此必須在分類上同時將此一政見分在：「地方建設」及「教育文化」這兩類。

但是，「爭取在桃園縣建設巨蛋球場」這樣的一個地方建設政見，還會有另一個問題。就是當我們依主管委員會職權分類時，要分出多少個主管委員會？雖然我們已將此一政見進一步分至「教育文化」中，但蓋巨蛋原則上也會涉及預算的爭取、人事的任用等有關財政委員會或經濟委員會所管轄的事務。為解決此一問題，研究團隊進一步訂定另一項規則為：**只需向外推「一層」至明顯主管其事的「立院委員會」**即可。畢竟所有的政策，幾乎都會有人事、預算的問題，若全部將其分類至這兩項類別中，未來在「經濟」及「財政」這兩類別上，勢必會出現數據被過度膨脹的現象。

至於如何分辨所謂「第一層主管其事機關相對應的立院委員會」，則如我們建立分類的方式，即以各立院委員會對其相對應的部會職權及委員會主管項目的說明為主。例如立法院的「經濟委員會」，其主要職掌事務為：經濟、農業、經濟建設、公平交易、能源、科技政策，以及經濟部、行政院農業委員會、行政院經濟建設委員會及行政院公平交易委員會掌理事項之議案。而「教育文化委員會」的職掌範圍則是：教育、文化政策；教育部、行政院文化建設委員會、國立故宮博物院、行政院新聞局、行政院青年輔導委員會、行政院體育委員會、中央研究院、行政院國家科學委員會、行政院原子能委員會掌理事項之議案。

由立法院對各委員會的功能定義，我們可以清楚地發現，「爭取在桃園縣建設巨蛋球場」這樣的地方建設政見，依研究團隊所訂的「第一層主管其事的機關」規則推論，分類必定是分在「教育文化」，而非「經濟」。

最後，本團隊登錄政見是以每一位候選人政見中的每一條目為登錄單元，但並非所有立委候選人提出的政見都是條目式，或是有明顯的政策分項。當候選人的政見是「非條列式」時，我們會強迫分條，分條的方式則參照下列的「計算準則」。而候選人的同一政見中，若是洋洋灑灑地列出多項性質重複度過高的政見，為避免數據被膨脹，我們也會因其屬性相同而僅做一次的政見登錄。以下為「議題分類」的計算準則、定義與案例的說明。

### (五) 計算準則

1. 地方建設與基礎建設，因為政見數量多，需向外推「一層」至主管其事的「中央部會」。例：「治水防洪兼觀光：爭取設置鎮南休閒區及全面整治筏子溪，發展精緻休閒農業，打造台中後花園」。除在議題登錄「觀光」及「基礎建設」各一次外，由於其主管單位為「經濟部」，同時亦登錄於議題「經濟」項中一次。
2. 若候選人所提政見為「非條列式」時：登錄時強迫分條，並計算未分條的政見比例。例如，第七屆台中市蔡錦隆：
  - (1) 「過去八年執政無能，意識型態治國，國家內耗，競爭力衰退，人民痛苦指數日漸高升。」這樣意義連接的句子，以句號或驚嘆號出現，列為一政見條目。
  - (2) 「改善民生、改革教育，才能重建正確價值觀；發展經濟、規劃都市，才能讓台中大繁榮！惟有微笑向前行，台中才有好未來！除了建設發展，更應有人文關懷；台中市向前行，台灣一定贏！」這樣地域方向近似的句子，列為同一政見條目。
  - (3) 「擬規劃爭取發展：汐止為電影數位文化園區，金山、萬里為休閒博奕特區，瑞芳、雙溪為觀光休旅特區，貢寮為國際音樂藝術展演區，平溪結合天燈與自然風光為飲食男女情慾解壓國際特區，以上須經住民過半數同意並經修法與配套規劃。」此一政見中，雖有重複的地方性「藝文、休閒、觀光」及外推一層「教育文化」的議題出現，但因其屬性相同，僅登錄各一次即可。若前後議題屬性差別較大，才做一次以上的登錄。

表 3 政見領域 (三)：議題類別說明

議題類別	
1. 由於地方建設與基礎建設的政見數量多，並且會涉及到不同的主管機關，需向外推「一層」至主管其事的「中央部會」。 2. 若候選人所提政見為「非條列式」時：登錄時強迫分條，並計算分條後的各政見比例。 3. 在同一條政見中，當內容中有重複性議題出現時，若其屬性過於雷同，則只需做一次的登錄即可，不需重複登錄。	
議題類別	定義、規則與範例
內政	服務 具有服務地方選民內涵者。 例：「爭取老人福利每里設置日間托老機構，老人津貼比照老農津貼以示公平。」
	藝文休閒觀光 非基礎及經濟建設。 例：休閒農場、現代新潮觀光夜市、公園、美術館、博物館、熱帶植物區
	地方 基礎建設 維持基本生活需求，不會馬上帶來地方經濟效益，必須在這基礎上再發展其他東西。 例：水電、新增道路、埋設光纖網路、捷運、強敏感地質的監測及山坡邊土石流防治，加速輕軌捷運，南港鐵路地下化、停車位（場）、廚餘回收、垃圾掩埋場、焚化爐。
	地方建設 會在短時間內為地方帶來經濟效率的建設。 例：遷機場、遷監獄、遷兵工廠、非水電道路基礎建設者、設勞動部在高雄、生物科技園區、大士林教育文化藝術學園、世界級經貿中心與企業總部、眷村改建。
	族群 與族群議題有關之立委政見。
	警政治安 與警政、治安、消費、打擊犯罪有關的立委政見。
	區域整合 （含國土規劃） 行政區域的升格、合併、改制等。
	政黨政治 與政黨運作、選舉有關之立委政見。
	其他 其他內政部管轄之事務。
	憲政議題 （修憲）
行政 行政機關權力的修正	
立法 立法機關權力的修正	
司法 司法機關權力的修正	
考試 考試機關權力的修正	
監察 監察機關權力的修正	
總統 總統權力的修正	
中央與地方 中央與地方權力的修正	
其他 其他憲法條文的修正	

表 3 政見領域（三）：議題類別說明（續）

議題類別	定義、規則與範例
國會改革	<p>定義 對於國會改革有具體建議之立委政見。</p> <p>實例 國會席次減半。</p>
外交國防	<p>定義 與行政院外交部及國防部、立法院外交委員會所職掌之事務有關之立委政見。</p> <p>實例 遷兵工廠。</p>
經濟	<p>定義 與行政院經濟部、立法院經濟委員會職掌事務有關之立委政見。</p> <p>實例 經濟犯罪、農漁林牧礦、有機農業、公營事業增加地區人事任用員額、增加就業機會、水庫、治水。</p>
財政	<p>定義 與行政院財政部、立法院財政委員會職掌事務有關之立委政見。</p> <p>實例 預算、金融、助學貸款、地方財政。</p>
教育文化	<p>定義 與行政院教育部、立法院教育及文化委員會職掌事務有關之立委政見。</p> <p>實例 體育活動、教改、加強弱勢族群教育經費。</p>
交通	<p>定義 與行政院交通部、立法院交通委員會職掌之事務有關的政見。</p> <p>實例 爭取高鐵站、公共工程、通訊傳播政策、遷機場。由於全國性觀光為交通部所執掌，亦列入此項。</p>
法制司法	<p>定義 1. 與行政院法務部、立法院司法及法制委員會之職掌事務有關的政見。 2. 候選人在政見有提出制訂、修訂的法案名稱。 3. 成立政府新的政府組織、政府單位時，所制訂的機關組織法案。 4. 各類證照制度。</p> <p>實例 證照制度、經濟犯罪、遷都、設勞動部在高雄、成立國家青年發展委員會</p>
福利	<p>定義 政見中指定給予全民、特定群體實質的補貼或免費的服務。</p> <p>實例 助學貸款、公營事業增加地區人事任用員額、增加就業機會。</p>
消費者保護	<p>定義 與消費者權益有關之政見。</p> <p>實例 協助解決消費者卡債問題、提昇消費者權益、落實消費者保護法。</p>

表 3 政見領域 (三)：議題類別說明 (續)

議題類別	定義、規則與範例
醫療衛生	<p>1. 與行政院衛生署職掌之職務有關之政見者列入此項。</p> <p>2. 當政見的主要內容為「醫療」或「衛生」行為時，才列入此項。例如「增取編列預算為婦女免費施打疫苗及健檢」的政見，主要內容為編列預算，而非醫療行為故不列入此項。</p> <p>實例 推動二代健保、提升醫療品質、整合三重醫療網。</p>
環境	<p>與環境保護、自然生態保護有關的政見。但若政見中強調的是單一地區的環境建設時，並不列入「環境」，而是列入「內政→地方→基礎建設」。</p> <p>實例 環保、有機農業</p>
其他	實不適於列入以上各議題類別者。如：目前尚未出現不適於以上各項議題分類條目者。

資料來源：本研究。

## 二、巢狀編碼

本資料庫的實際登錄過程，為了與文本探勘的技術做結合，也同時對立委選舉的公報政見進行「巢狀編碼」。

所謂資料探勘 (Data-mining) 或文本探勘技術，都是利用資訊科技，以電腦軟體管理、組織、分析大量數位文件的資訊技術，近幾年來逐漸被用在社會科學領域當中。而文本探勘更特別用以處理文字、詞彙為主 (Tsoumakas, Katakis, and Vlahavas 2010)。

為了配合文本探勘技術的運用，研究團隊對立委政見進行「巢狀編碼」，試圖教導電腦學習人腦對於政見的分類方式。原則上，即是以中括號標示研究團隊對立委政見中什麼樣的字彙會起動「分類的念頭」，然後再以小括號標示「該字彙被分類到哪個資料庫的分類標的」。例如：「推動族群平等法，促進族群融合」，這樣的一個政見，依上節研究團隊對政見的分類及定義，會被分在「全國性（政見地域方向）、政策意涵（政見性質）、內政族群（議題類別）、法制司法（議題類別）」。而這樣的分類，反映在「巢狀編碼」則是：「〔推動〔族群平等法〕 (4-9)，促進〔族群融合〕 (4-1-2)。〕 (1-1,2-3)」。<sup>3</sup> 每一個數字代表

<sup>3</sup> 各編碼如 4-1-2, 1-2, 2-3 等，各有意義。第一層有 1 到 4 之編號，代表政見分類的四大面向（即 1 地域方向、2 性質、3 對象及 4 議題）；第二層為每一面向下的次分類，如地域方向下有全國、單一、跨區、兩岸，及國際五個次分類，依序編號為 1 到 5。所以政見地域方向上出現為「單一地方」，就會登錄為「1-2」；第三層則是次分類下的細分類，如議題（編碼為 4），下有內政外交等次分類（編碼為 4-1 到 4-14）；內政次分類下又有細分類如地方等，就再往下編為：4-1-1；地方細分類下再有「服務、藝文休閒」等子分類，就編為：4-1-1-1。依此類推，全部編碼層級最多即為四層。

相對應的都是研究團隊對資料庫的分類項目，例如 (4-9) 即是「法制司法」，表示編碼的人員看到「族群平等法」這樣的政見字彙，會將其歸類到「法制司法」這個議題項目當中。最外圍的中括號，則是表示編碼人員看到整個政見：「推動族群平等法，促進族群融合」，這些字彙，會反映這一政見的地域方向是全國 (1-1)，性質是政策意涵 (2-3)。

再以另一政見：「推動身心障礙者終身教養，並在南投縣成立啓智學校」為例。此一政見在資料庫的分類位置會是「單一地方（政見地域方向）、有政策意涵（政見性質）、殘障（政見對象）、教育文化（議題類別）」。**反映在「巢狀編碼」則是：「〔推動〔身心障礙者〕 (3-21) 終身教養，並在南投縣〔成立啓智學校〕 (4-7)。〕 (1-2,2-3)」。**表示編碼人員會將「身心障礙者」一詞分類為殘障 (3-21) 對象，將「成立啓智學校」分類為教育文化議題類別 (4-7)。最外圍的中括弧則是編碼人員認為「推動身心障礙者終身教養，並在南投縣成立啓智學校」，此一政見是單一地方方向 (1-2)，及有政策意涵 (2-3) 的。

巢狀編碼的優點除了教導電腦學習人腦的思考方式，以及在日後對大量的政見文字檔案進行快速有效率的分類處理外，亦可將立委政見做明確的分類標示，也代表團隊任一成員在對政見進行分類時，是可以被檢驗其分類標準的。當研究人員對大量複雜的政見文本做檢閱及分類的工作時，難免會對一些語意不清、文字分項不明的政見產生解讀不一，或無法判別分類的問題。當我們明確做出分類標示時，代表團隊成員若是對同樣或相近的文字做出不同分類的時候，可以輕易地找出問題的根源，並且讓所有成員都可以有討論的共同平台，而不會產生找不出問題的所在，或是無法討論分類標準的現象。這樣一個「巢狀編碼」的動作，也等同強化了此一「政見資料庫」建置的信度。

### 三、信度檢核

建構資料庫是許多研究者賴以檢證理論的步驟之一，良好的資料庫除了分類與建構方式有理論的支撐外，研究團隊成員登錄資料的標準是否統一、是否有依照資料庫的建構規則與分類定義進行編碼、以及整體登錄數據資料的正確程度，都會影響到資料庫的信度。<sup>4</sup>

本研究「第六及第七屆立委候選人的政見資料庫」，為確保資料庫信度，曾採取下列幾個步驟來檢驗：

(一) 模擬分類：在 98 年的 4 月至 9 月間，研究團隊每週固定舉行會議，討論團隊

<sup>4</sup> 本文不談論「效度」的問題，因效度主要指概念定義 (conceptual definition) 及操作化定義 (operational definition) 間是否契合。而本文主要是討論對實際政見的資料庫建構，重要的是分類的方式是否有理論基礎、人員對政見的編碼是否有一致性，因此本文在此只介紹研究團隊保持資料庫信度的方式。

成員每週對所分配到的候選人政見的模擬分類狀況。這段期間中，除了建立有共識的分類標準外，四個多月的模擬分類，也讓每個團隊成員瞭解所負責登錄區域立委政見的性質，大大降低了日後實際分類出錯的機率。特別是許多資料庫的規則是需要經過多次分類才能夠熟悉的，如地方建設與基礎建設，因為政見數量多，「需向外推一層至主管其事的中央部會」的規則，必須經過不斷地討論，以及每個編碼者的反覆揣摩，才能夠熟稔每個有政策意涵的政見應分類到哪一個議題類別。

- (二) 政見編碼手冊 (coding manual) 的編製及 codebook：為了統一規則與編碼有所依據，本團隊擬出政見編碼手冊與 codebook，讓每個團隊成員都有編碼的依循標準，並可隨時查閱編碼規則。前者政見編碼手冊（參見附錄一範例），記錄了每一個政見分類的詳細定義及範例，以及每一次團隊開會的案例討論結果、各部會及立法院各委員會的功能職掌，以及在第六屆及第七屆的立委政見中會出現的時事性議題介紹，例如新十大建設的定義及內容。後者 codebook 則是以 excel 檔案的形式（參見附錄二 excel codebook 範例），每一欄位都有相對應的分類及編碼方式、分類定義註解。

政見編碼手冊與 excel codebook 的建立，確保了每位政見分類的編碼者都有共同依循的準則及分類方式，其中 excel codebook 每個分類標的，也因此有了相對應的代碼。在「附錄二」中即可清楚見到，若一政見的地域方向為「全國性」，則其代碼即為 (1-1)；若一政見性質為政治口號，則其代碼為 (2-2)。這樣的編碼方式也提供了巢狀編碼編號的來源。

1. 巢狀編碼：如上所述，巢狀編碼雖然最初的目的是要教導電腦軟體，人對政見的分類標準思維是如何運作的，以利日後應用電腦軟體來處理大量的政見文字資訊。但是這樣的編碼過程，同時也建立了可供日後互相檢核的資訊。每一個編碼者可以輕易的透過巢狀編碼，去檢視另一個編碼者的分類思維邏輯是否與團隊的標準一致。
2. 交互檢驗：98年9月整個資料庫的建構分類及規則正式穩定下來，不再有大程度的調整之後，研究團隊正式開始對第六及第七屆立委政見資料庫進行分類及編碼。在編碼的同時，我們將編碼人員兩兩分成一組，並要求每一個人員在編碼完成後，交由同組另一成員檢驗。小組內兩人交互檢驗時，互相對彼此的編碼進行討論及除錯。小組的兩位成員編碼都完成之後，再進行一次小組與小組間的交叉檢驗，同樣的步驟再度在小組間進行一次，檢驗完成後，提出小組與小組之間的思維誤差，共同討論及除錯。
3. 最終檢驗：98年10月起，完成上述步驟的立委政見資料庫編碼，彙整至整個資料

庫，由小組成員（李承訓及陳月卿）進行電腦輸入及分析。在輸入電腦資料庫的同時，此一小組再度進行最終的除錯，特別是巢狀編碼的括符位置錯誤，或是編碼欄位跳格等細微部分，再度進行檢驗。由於此一步驟同時利用了電腦進行判讀，因此一些肉眼較不易辨視出來的錯誤，都在此時由輸入小組進行調整。

誠然，這樣的資料庫建構，很多時候涉及了編碼人員的個人判斷以及對資料庫分類定義與規則的掌握程度，每位編碼人員還是會有所差異。但是團隊利用此五項檢驗步驟，盡可能的將個人判斷與掌握程度的差異性降到最低，也將跳格或編碼位置出錯的機率盡可能的減少。具體的信度數據，在初階兩人一組編碼人員的組內「交互檢驗」時，編碼不一致比率，平均達到 28%，到「最終檢驗」前，已降至平均 7% 左右，而這些剩餘的不一致，最終由兩位博士生做逐一的討論及調整，務必求新舊手的判讀登錄完全一致。這些檢證步驟可說將本資料庫的內部信度達至百分百。

## 肆、資料探勘技術應用——多標籤文件分類法

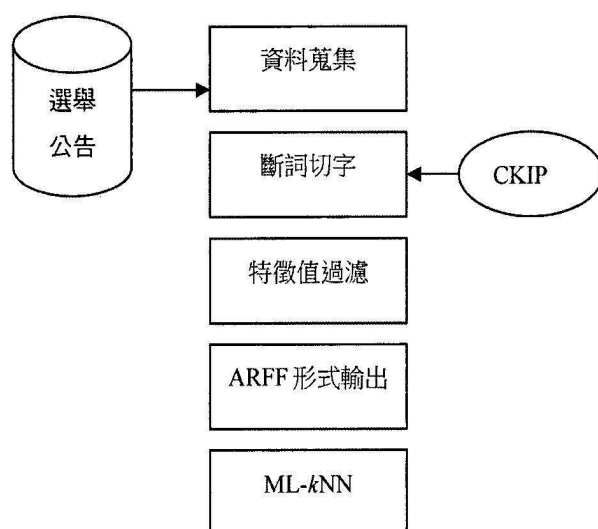
### 一、系統機制

雖然競選承諾已經初步藉由人工編碼分類為四大面向，但文本探勘技術目前並沒有進展到能夠同時將四種面向分類的地步。本研究團隊決定選擇競選承諾之議題面向做為測試人腦思考方式與電腦學習成果一致性的開始。<sup>5</sup>

由於競選承諾的複雜性，每個承諾可能不僅屬於單一類別。為了處理這個問題，多標籤文件分類 (multi-label classification) 在近來的文本分類 (text categorization) 領域越來越受到重視。多標籤文件分類法主要有兩大研究方式 (Tsoumakas and Katakis 2007)：問題轉換法 (problem transformation method) 以及算法改造法 (algorithm adaptation)。前者將多標籤文件分類工作轉換成單一或多個單一標籤分類、回歸分析或標籤排序工作；而後者則是延伸既有的學習算法，直接解決多重標籤資料 (Katakis, Tsoumakas, and Vlahavas 2008)。

為了漸進地達成政治選舉承諾分類的目標，本研究採取  $k$  最鄰近法 ( $k$ -Nearest neighbor approach,  $k$ NN)，加上多標籤文件分類法，稱為多標籤  $k$  最鄰近法（以下簡稱 ML- $k$ NN）做為多標籤文件分類方式 (Zhang and Zhou 2005)。ML- $k$ NN 的優點在於研究者不需要反覆重新建立訓練模式，且前置作業較不複雜，因此容易採行。整體的資訊探勘流程和如圖 2。

<sup>5</sup> 相較於其他三個面向：地域方向，性質與對象，競選承諾的議題面向最難藉由電腦軟體分類，因為議題面向經常包含多標籤分類內涵，因此本研究團隊於一開始便採用 ML- $k$ NN 這個新方法。



資料來源：本研究。

說明：「特徵值過濾」是以文本的內容特徵來進行過濾。「ARFF(Attribute-relation file format)」為一種數據儲存格式，例如基於 JAVA 的 WEKA 資料探勘軟體，即以此種格式儲存資料；本研究所使用的多標籤分類工具，即是基於 WEKA 架構之應用軟體。

圖 2 資訊探勘流程圖

## 二、前置作業

在多標籤文件分類過程中，首先，我們使用中央研究院中文斷詞系統 (Chinese Knowledge and Information Processing, CKIP) 將文章中的字句切割出具有意義的詞彙並標註詞性，<sup>6</sup> 且利用 N-gram 模式的概念將所有連續出現之名詞組合在一起，產生新的詞彙。我們僅採用名詞此一詞性做為實驗對象，並將不具代表性之詞彙過濾，僅留下重要之名詞詞彙。因此我們採取 Weka 這個語詞特徵篩選工具 (Hall et al. 2009)，透過卡方檢定僅保留具代表性之屬性。最後，為了訓練分類器，我們將資料轉換成 ARFF(Hall et al. 2009) 形式。

### (一) 多標籤 $k$ 最鄰近法 (ML- $k$ NN)

前置作業之後，透過 Mulan 處理 ML- $k$ NN 算法 (Tsoumakas, Katakis, and Vlahavas 2010; Zhang and Zhou 2007)。這種演算法，如前所述，是結合多標籤分類法及  $k$  最鄰近法，前者特別適用於結構複雜，如有情感意涵的文本；後者則是最常見的分群方法，但通常需要較大量的文本，所以「政見」投入的資料量必須盡量大，兩者結合，也是本文較大膽的應用嘗試。

<sup>6</sup> 中文斷詞方法仍有許多待突破之處，如簡稱→內湖捷 (內湖捷運)，中研院系統即無法判讀；改進之方法，或為建立各領域的「領域詞典」，協助電腦識別，但這就須長期的跨領域合作。

## (二) 評估方式

根據 Zhang 與 Zhou(2006)，多標籤文件分類最常用的方式是：漢明損失 (Hamming loss)、1 錯誤率 (one-error)、覆蓋率 (coverage)、排序損失 (ranking loss)、平均精度 (average precision)。<sup>7</sup>

## (三) 結果

我們將第六屆競選承諾議題面向分類做為訓練資料，而將第七屆競選承諾做為測試資料。Binary relevance learning(BR) 和 Label power set learning(LP) 是兩個經常被使用的問題轉換方法。每種量度的最佳結果以粗體字顯示如下。ML-*k*NN 法對照比較 BR 和 LP 方法的表現結果。從表 4 我們可以發現 ML-*k*NN 法包含最低的漢明損失、1 錯誤率、覆蓋率、排序損失和最高的平均精度，這指出 ML-*k*NN 實質上表現比其他算法更好。

表 4 三種算法表現結果評估

Measurement	漢明損失	1 錯誤率	覆蓋率	排序損失	平均精度
ML- <i>k</i> NN	<b>0.0255</b>	<b>0.8637</b>	<b>2.4565</b>	<b>0.1291</b>	<b>0.4403</b>
BR(SVM)	0.0257	0.9147	6.3044	0.3564	0.2445
LP(SVM)	0.0465	1.0000	9.2311	0.5497	0.1040

資料來源：本研究。

此結果提示我們可以用 ML-*k*NN 法進行包含「對象」面向的第七屆競選承諾之分類，因為既定的一個競選承諾中可能包含一個以上標籤（或對象）。而 Tsoumakas 與 Katakis (2007) 認為每個競選承諾在「地域方向」和「性質」面向上僅限於單一標籤而應更易於分類。因此本文選擇不列出這些面向應用的技術性細節。

目前人工與資訊工具的分類比對結果，資訊工具 (ML-*k*NN) 的準確率，尚未達百分百精準，還需要後續更大量的資料投入訓練；或者未來可先從較簡易的「單一標籤」分類項目，如「地域方向」著手，來從事自動化後的人工確認。

<sup>7</sup> 1 錯誤率：我們可以將分類器定義為  $H: X \rightarrow Y$ ，將單一標籤透過設定  $H(x) = \operatorname{argmax}_{l \in Y} f(x, l)$  指定給例  $x$  (instance  $x$ )。而 1 錯誤率評估最高標籤不在實際分類標籤中的次數。

漢明損失：漢明損失將預測誤差（預測不正確的標籤）以及失誤（未能預測的標籤）納入考慮。我們考慮學習函數  $f: X \rightarrow 2^Y$ ， $f$  測試資料 (test set) 的漢明損失  $T, HL(f)$  即定義為：  
 $\frac{1}{km} \sum_{i=1}^m \sum_{j=1}^k \delta(l \in f(x_i) \wedge l \notin Y_j) + \delta(l \notin f(x_i) \wedge l \in Y_j)$ 。

覆蓋率：有時候我們要以樣本所有可能的標籤評估一個系統的表現，此時僅有最高標籤便不足夠。覆蓋率乃是樣本  $x$  所有可能標籤的平均距離。

平均精度：平均精度可測量預測標籤的平均精確度。測量在特定標籤  $l \in Y_i$  以上所有標籤但卻實際是在  $Y_i$  的數值。

## 伍、結論

本文試圖使用中央選舉委員會所公布的選舉公告，顯示第六屆與第七屆立法委員候選人承諾資料庫建立過程。本文主要目的在於發掘各競選承諾的多面向性質，藉以推動政治科學領域研究工具的進展。本文嘗試說明如何結合人工編碼與文本探勘法對上述競選承諾進行分類。

本文首先說明第六屆與第七屆立委候選人競選承諾人工編碼完成過程。基於競選承諾的多面向性質，人工編碼首次將競選承諾分類為四個基本面向：地域方向、性質、對象與議題。本文一開始便選擇難處理的議題面向做為測試同樣具有多面向的電腦軟體所表現的人工智慧。

我們利用 ML-*k*NN 文件探勘法對個別競選承諾的議題進行分類。本文證明這種多標籤分類算法在與人腦思考合作上較其他方法佳。因為競選承諾資料庫最困難的部分已透過資訊通訊科技 (Information communication technology, ICT) 解決，本研究團隊有信心在未來，至少在政治科學的文件分析或分類上能夠結合 ICT 技術。然而，是否 ML-*k*NN 法與其他算法可以自動應用於之後的第八屆立法委員競選承諾，仍需要更多解釋與實驗。本文截至目前為止尚未提供保證。但本研究團隊將持續以結合 ICT 工具的分類機制做為未來所要達成的長期目標。近期內，則希望做到半自動化分類，也就是先用文字探勘技術做自動分類，再用人工進行校正，尤其是政見的「地域方向」或「對象」這樣較具體的面向，可以是先著手的地方。

除了上述方法之外，第六屆、第七屆台灣立法委員競選承諾資料庫透過人工編碼建立將帶來諸多研究上的用處。如本文前面幾節所提，候選人競選承諾結構與內容本身、其與選舉結果之關係，以及與其立法表現之間的關係等，皆顯示透過此資料庫可以進行的研究主題。再者，諸如選舉制度變化、選區特色以及立委背景的影響等皆可透過此資料庫做研究探討。

目前初步的探究，即可看出一些或可與理論對話的趨勢，如第七屆的立委較第六屆明顯的多提出單一地方方向的政見，但口號性政見（含非政治及政治性）又大量少於第六屆。如以政黨來區分，國民黨比較重視兩岸議題，民進黨則是國際議題；而國民黨當選者較注重個人特質，民進黨當選者則需要較多的政治性口號。換句話說，政見的結構特質似與政黨有關，也與候選人的輸贏有所關聯。這些發現的背後值得玩味，也等待研究者的深入探索。

綜上所述，在台灣，競選承諾做為官方文件，廣泛在選民間流傳，對於民主運作具有象徵性與實質性雙重意義。即使仍有許多台灣人可能並未注意到這些競選承諾，研究者仍

不應忽略這些文件的存在，以及這些文件做為承載選民及其代表之間的「契約」性質。研究者若能努力發掘一個政治人物在這些官方文件中的承諾，和其上任後工作上所實踐的，兩者之間的連續與不連續性，則該研究可以對政治人物課責，進而達成深化民主的目的。因此，促使台灣民主更具有可課責性，也是建立資料庫本身的目標。這個資料庫歡迎任何對於上述相同努力有興趣的人士，在未來加以發揮運用。

\* \* \*

投稿日期：2011.12.15；修改日期：2012.05.22；接受日期：2012.06.15

## 附錄一、立委政見編碼手冊範例

各欄位定義：

1. **方向**：指立委所提的政見，其內容涉及的地域範圍為何，並依地域範圍分為：
  - (1) 全國性，政見內容屬全國普遍適用者，如：修憲、募兵等。
  - (2) 單一地方，政見內容僅適用於單一地方者，如：建設信義區、南港鐵路地下化等。  
規則：單一地區的對象是「全民」時，不需 coding 對象。
  - (3) 跨區性，政見內容橫跨不同選區者，如：基隆河之整治、中橫公路之修繕。
  - (4) 兩岸，政見內容特別提及「台灣與大陸之各種形式的交流」者，如：反對兩岸軍備競賽。  
規則：當政見提及「兩岸和平」時，其必定涉及「議題→外交國防」。
  - (5) 國際，指政見內容為參與國際組織或其他與外國交流之事務。如：參加國際組織：聯合國、WHO、WHA、海洋台灣等。
  
2. **性質**：指該政見的屬性為何，區分如下：
  - (1) 非政治性口號：民生或經濟議題，但是沒有比較具體實踐的方法或目標，有說法沒做法。例「揮別苦日子、再創好生活」、「貼近人民心聲、效率責任自許」。  
規則：當政見為非政治性口號時，不列入議題類別，但可以有對象。
  - (2) 政治口號：難以在公共政策層次上釐清，必須進行政治動員的政見（含：涉及國家領土主權的認定）。例：黨產、貪腐、修憲、討黨產、進入聯合國、臺灣國、去中國、支持軍購保衛台灣、反軍購、台灣主體、族群平等、原住民地位、國家（領土、土地）資源資源共管入憲。

## 附錄二、Excel codebook 範例

C	D	E	1. 方向					2. 性質					3. 對象									
			1. 全國性	2. 單一地方	3. 跨區性	4. 兩岸	5. 國際	1. 非政治性口號	2. 政治口號	3. 政策	4. 個人特質	1. 全民	2. 軍公教	3. 台商	4. 老人	5. 婦女	6. 原住民	7. 族群	8. 外籍人士(含外籍配偶)	9. 學生	10. 中壯年	11. 青少年
立委候選人	黨籍	政見順序 0. 當選 1. 未當選																				
<b>USER:</b> 定義：難以在公共政策層次上釐清，必須進行政治動員的政見 含：涉及國家領土主權的認定。 例：黨產、貪腐、修憲、討黨產、進入聯合國、臺灣國、去中國、支持軍購保衛台灣、反軍購、台灣主體、族群平等、原住民地位、國家(領土、土地)資源共管入憲、支持陳水扁總統所提出的政見。																						

## 參考文獻

### I. 中文部分

王鼎銘，2003，〈政策認同下的投票效用與選擇：空間投票理論在不同選舉制度間的比較〉，《選舉研究》，10(1): 171-206。

(Wang, Ding-ming. 2003. "Voting Utility and Choice Decision in 2001 Election: The Application of Spatial Voting Theory in Different Electoral Systems." *Journal of Electoral Studies* 10(1): 171-206.)

林季福，2003，〈九十年立法委員選舉候選人體育運動政見之分析研究〉，《台東師院學報》，92(14): 405-420。

(Lin, Chi-fu. 2003. "The Analysis of the Political Viewpoints on Physical Education and Sports Given by the Candidates of 2001 Taiwan Area Legislators' Election." *Journal of National Taitung Teachers College* 92(14): 405-420.)

林東穎，2010，〈立法委員競選政見的提出與選後立法表現的連結性——以第六、七屆高雄市立法委員為例〉，中山大學政治學研究所碩士學位論文。

(Lin, Tung-ying. 2010. "The Campaign Promises of Taiwanese Legislators and Their Performances in the Legislative Yuan: The Case of the 6th and 7th Term Legislators from Kaohsiung City." Master's thesis. National Sun Yat-Sen University.)

胡佛，1986，〈選民的政見取向：結構、類型與運作分析〉，《社會科學論叢》，34: 111-143。

(Hu, Fu. 1986. "Voters' Issue Orientations: An Analysis of Structures, Typologies and Operations." *Journal of Social Science* 34: 111-143.)

徐火炎，1991，〈政黨認同與投票抉擇：台灣地區選民的政黨印象、偏好與黨派投票行為之分析〉，《人文及社會科學集刊》，4(1): 1-57。

(Shyu, Hou-yan. 1991. "Party Identification and the Vote Choice: An Analysis of Electorate's Party Images, Partisan Preferences and Voting Behavior." *Journal of Social Sciences and Philosophy* 4(1): 1-57.)

-----，1993，〈選民的政黨政治價值取向、政黨認同與黨派投票抉擇：第二屆國大代表選舉選民的投票行為分析〉，《國家科學委員會研究彙刊：人文及社會科學》，3(2): 144-166。

(-----, 1993. "Voters' Value Orientations toward Party Politics, Party Identification and the Vote Choice of Parties: An Analysis of Voters' Voting Behaviors during the Second National

Assembly Election.” *Proceedings of the National Science Council. Part C, Humanities and Social Sciences* 3(2): 144-166.)

陳永峰，2005，〈政見在選戰中扮演的角色之研究——以 2004 年立委選舉為例〉，逢甲大學公共政策研究所碩士學位論文。

(Chen, Yung-fong. 2005. “The Role of Platforms in Campaigns: A Case Study of 2004 Legislator Election in Taiwan.” Master’s thesis. Feng Chia University.)

黃國敏，2009，〈縣市長選舉政見的實證分析：苗栗縣個案分析〉，中華大學行政管理學系第四屆「地方政府與公共管理國際學術研討會」，10 月 23 日，新竹：中華大學。

(Huang, Kuo-min. 2009. “An Empirical Analysis of Campaign Promises Delivered by County and City Mayors: A Case Study on Miao-Li County.” Presented at the Department of Public Administration, 4th International Conference of Local Government and Public Administration in Chung Hua University, Hsin Chu.)

廖益興，2005，〈競選行為效應的檢視〉，《中華行政學報》，2: 95-102。

(Liao, I-hsing. 2005. “Review the Campaign Effects.” *The Journal of Chinese Public Administration* 2: 95-102.)

劉從葦，2006，〈台灣政黨的政策位置：非介入式與介入式測量的比較研究〉，《台灣政治學刊》，10(2): 3-62。

(Liu, Tsung-wei. 2006. “Policy Positions of Political Parties in Taiwan: A Comparative Study of Unobtrusive and Obtrusive Measurements.” *Taiwan Political Science Review* 10(2): 3-62.)

魏中仁、楊志良，1997，〈立委選舉之醫療保健政見分析，1965-1995〉，《中華公共衛生雜誌》，16(5): 435-442。

(Wey, Chung-jen, and Chih-liang Yaung. 1997. “Health Care Political Views Showed in Legislator Election Campaigns, 1969-1995.” *Chinese Journal of Public Health* 16(5): 435-442.)

## II. 外文部分

Arnold, R. Douglas. 1990. *The Logic of Congressional Action*. New Haven: Yale University Press.

-----, 2004. *Congress, the Press, and Political Accountability*. New Jersey: Princeton University Press.

Budge, Ian. 2001. “Validating Party Policy Placements.” *British Journal of Political Science* 31: 210-223.

- Budge, Ian, and Dennis J. Farlie. 1983. *Explaining and Predicting Elections: Issue Effects and Party Strategies in Twenty-Three Democracies*. London: George Allen and Unwin.
- Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper & Row Publishers.
- Elling, Richard C. 1979. "State Party Platforms and State Legislative Performance: A Comparative Analysis." *American Journal of Political Science* 23: 383-405.
- Fine, Terri S. 1994. "Interest Groups and the Framing of the 1988 Democratic and Republican Party Platforms." *Polity* 26: 517-530.
- Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. 2009. "The WEKA Data Mining Software: An Update." *ACM SIGKDD Explorations Newsletter* 11(1): 10-18.
- Katakis, I., G. Tsoumakas, and I. Vlahavas. 2008. "Multilabel Text Classification for Automated Tag Suggestion." Presented at the ECML/PKDD 2008, Antwerp.
- Monroe, Alan D. 1983. "American Party Platforms and Public Opinion." *American Journal of Political Science* 27: 27-42.
- Pomper, Gerald M. 1968. "If Elected, I Promise: American Party Platforms." *Midwest Journal of Political Science* 11: 318-352.
- , 1980. *Elections in America: Control and Influence in Democratic Politics* (Rev. ed.). New York: Longman.
- Roberson, David. 1976. *A Theory of Party Competition*. London: John Wiley and Sons.
- Sulkin, Tracy. 2012. *Issue Politics in Congress*. Cambridge: Cambridge University Press.
- Tsoumakas, G., and I. Katakis. 2007. "Multi-Label Classification: An Overview." *International Journal of Data Warehousing and Mining* 3(3): 1-13.
- Tsoumakas, G., I. Katakis, and I. Vlahavas. 2010. "Mining Multi-Label Data." In *Data Mining and Knowledge Discovery Handbook*, eds. Oded Maimon and Lior Rokach. New York: Springer.
- Zhang, M. L., and Z. H. Zhou. 2005. "A k-Nearest Neighbor Based Algorithm for Multi-Label Classification." Presented at Granular Computing, 2005 IEEE International Conference, Beijing.
- , 2006. "Multilabel Neural Networks with Applications to Functional Genomics and Text Categorization." *IEEE Transactions on Knowledge and Data Engineering* 18(10): 1338-1351.
- , 2007. "ML-KNN: A Lazy Learning Approach to Multi-Label Learning." *Pattern Recognition* 40(7): 2038-2048.

# The Establishment of Taiwanese Legislators’ Campaign Promise Database

Da-chi Liao<sup>\*</sup> · Fu-ren Li<sup>\*\*</sup> · Yu-ci Huang<sup>\*\*\*</sup> ·  
Zi-yu Liu<sup>\*\*\*\*</sup> · Cheng-xun Lee<sup>\*\*\*\*\*</sup>

## Abstract

In Taiwan’s legislative elections, candidates running for office are required to submit campaign promises to be published in the Electoral Bulletin. This research considers that campaign promises published in Electoral Bulletins are the official commitments of the candidates and should not be disregarded as insignificant and unworthy of research. Moreover, the connection between the candidates’ campaign promises and their performances after being elected should be studied in order to monitor their “accountability” to voters. It is our hope that more emphasis is placed on “campaign promises” printed in Electoral Bulletins, which would help to create a positive influence on the creation of a democracy.

In order to achieve the above stated goal, this paper has established the “Sixth and Seventh Legislative Campaign Promise Database” based on campaign promises by candidates during Taiwan’s sixth and seventh term legislative elections. Hopefully this database would facilitate the research on issues related to the connection between campaign promises and the legislators’ on-job performance after being elected.

This paper will mainly cover the establishment process of this database,

---

<sup>\*</sup> Director and Professor, Institute of Political Science, National Sun Yat-sen University.

<sup>\*\*</sup> Director and Professor, Institute of Service Science, National Tsing Hua University.

<sup>\*\*\*</sup> MA, Institute of Information Management, National Tsing Hua University.

<sup>\*\*\*\*</sup> Ph.D Student, Institute of Political Science, National Sun Yat-sen University.

<sup>\*\*\*\*\*</sup> Ph.D Student, Institute of Political Science, National Sun Yat-sen University.

including the manual coding method and the latest information technology used for data mining, namely the multi-label classification method. This paper also attempts to compare and contrast these two methods and hopes to bring in the most advanced information technology to conduct a certain level of automated classification for campaign promises printed in Electoral Bulletins in every election.

Keywords: Legislative Yuan, legislators, campaign promise database, data mining, multi-label classification