

行政院國家科學委員會專題研究計畫 成果報告

建立代理者溝通機制：語意表示法與交換問題之研究

計畫類別：個別型計畫

計畫編號：NSC91-2213-E-004-013-

執行期間：91年08月01日至92年07月31日

執行單位：國立政治大學資訊科學系

計畫主持人：劉昭麟

計畫參與人員：張正宗，何君豪等

報告類型：精簡報告

處理方式：本計畫可公開查詢

中 華 民 國 92 年 10 月 21 日

建立代理者溝通機制：語意表示法與交換問題之研究

劉昭麟

國立政治大學資訊科學系

chaolin@nccu.edu.tw

前言

本研究計畫原始提案原本為一個整合型計畫中的子計畫，不過因為整合之母計畫未獲核准，原本的子計畫也未全數通過，因此子計畫間的合作機會受到很大的影響，有許多原本規劃合作的工作不能夠執行，所以本計畫執行的目標和成果跟原本所提的研究計畫有所出入。

本研究轉而從事研究利用人工智慧技術分析法院判例的文書的文義，然後歸納出審判規則，並利用這一些規則用於未來案件的審判工作上。雖然我們的工作脫離了代理者溝通的問題，不過仍舊維持在研究語言的意義的方向上。過去一年當中，我們從無到有積極工作，因此能在計畫執行期間將成果呈現在兩個具有中上水準的國際學術會議和一篇國內的學術會議中。

成果摘要

✓ 研究標的的選擇：簡易法庭案件

政治大學法學院陳起行教授的研究發現，我國學者從事法資訊學的研究明顯偏少。基於資料電子化的趨勢，我們利用這一年的研究經費從事法資訊學的先導研究，建構了一套可以輔助簡易法庭判決的系統雛形。簡易法庭審理的案件屬於法院所審理案件中案情比較單純的訴訟案件。在相關的起訴狀中，大都只要幾百個字就足以描述被告的行為，並據以決定相對應之處罰。相對於一般刑事和民事案件，簡易法庭的案件要容易的多，民事案件相關文書中對被告行為的描述經常是非常的長，牽涉到許多細節，要利用軟體系統處理，極可能必須依賴許多的背景知識，對一個先導性的研究的挑戰性過大，因此經過一番嘗試和思考，本研究案從簡易法庭的案件著手。此外，有一些簡易法庭案件會牽涉到多位涉案人，相對的案件內容的描述也相當複雜，因此我們還暫時限制我們所處理的案件只有一個被告，同時被告最後只有以一個理由被起訴（以下簡稱一人一罪案件）。我們已經在後續的研究中，開始嘗試比簡易法庭所處理的案件更複雜的多人多罪的案件。

✓ 研究背景

國外，尤其是歐洲國家，對於法資訊學的研究已經從事多年。相關的研究在國際人工智慧與法資訊學學會的期刊和歷年的研討會論文集可以找到許多的資料(<http://www.iaail.org>)。前人的研究，有就法學知識的表達方法的研究，有

對法學領域的自動推論的研究，有其他許多特殊應用問題的研究。本研究利用最常見於相關文獻的知識表達法和推論機制 - 案例式推論系統 (case-based reasoning, CBR)，從事簡易法庭中案件起訴理由（以下簡稱案由）和適用法條的分類工作。

由相關文獻中可知，過去多年雖然 CBR 被相關學者廣泛採用，但是有許多系統是以人工的方式建立案例資料庫 (case databases)。這樣的做法雖然可以讓系統有較好的表現 (performance)，但是系統建置工作的人力和時間代價其實很高。因此有學者呼籲研究工作應該朝向自動化建立案例資料庫，才能真正的讓 CBR 技術進入實際應用中。有鑑於此，本研究提出以機器學習 (machine learning) 的方法，嘗試讓程式自動建立案例資料庫。實驗的結果顯示，雖然程式不能夠完全取代人工，但是我們所提供的演算法卻有可能可以輔助和加速人工建制案例資料庫。

✓ 研究方法

我們蒐集簡易法庭中一人一罪的判例的起訴書和判決書作為訓練和測試語料。目前，我們擁有三季共約一千七百個判例。其中第一季約五百筆判例用作訓練語料，其他兩季的資料則用作測試語料。¹

我們的訓練步驟是以第一季的資料做為輸入資料，針對每一個過去的判例，以一定的機制產生精簡的判例資料檔案，然後將這一些檔案儲存在案例資料庫中。在測試階段，我們用 k-Nearest-Neighbors 的技術，計算測試案例跟過去的判例的相似程度，然後篩選出若干判例來決定測試案件的案由和適用法條。

顯然的，若是將所有過去判例都產生一筆判例資料檔，將來不管是資料儲存或是應用都會付出計算量和計算時間的代價。因此我們進一步設計了一些可以合併和簡化判例資料的方法，將相似的案件的判例資料合併，並且將判例資料檔中的資料去蕪存菁，以改進計算上的效率。我們的訓練和測試方式的細節請參考本摘要所附的實際論文，下面僅提供一些方法的概述。

如果我們的訓練判例中有很相似的案例，則我們將相關文件斷詞之後所擷取出來的詞就會有相當的相似性。所謂相似性包含詞本身的雷同和詞的順序的雷同。由於事件發生的順序是違法行為的一個極重要因素，因此我們在分析案件類似性上也需要考量到詞的順序，事實上這也是 case-based reasoning 的一個重要特徵。依據過去判例所使用的辭彙的相似性，我們的程式判斷兩個判例是否相似，如果是相似案例，則以兩者共同的部分取代原本兩個案例。

如果有兩個案例，其中一個案例包含了另一個案例相當比例的辭彙，我們則大膽的將這個被包含的案例自案例資料庫中刪除。上述這兩個基本原則，可以

¹ 一般而言，訓練語料會多於測試語料。訓練語料多，可以提高系統的表現。我們這種異於尋常的做法是因為我們的語料來自法院，為了保護個人隱私，資料的取得其實不容易，因此有很長的一段時間，我們只有兩季的資料，所以用一季的資料作為訓練，另一季的資料用於測試。得到第三季判例資料之後，我們就沿用以第一季資料所訓練出來的系統。

幫助我們減少案例資料庫的案件的量。

此外，我們還以個別的辭彙出現在判例中的頻率決定詞的重要性。這一個做法有點類似傳統資訊檢索 (information retrieval) 技術中的 TFIDF (term frequency inverse document frequency) 的技術，但是不盡相同。我們從既有的 case 中刪除出現頻率不及所選定的一個最低頻率而且同時出現在兩種不同案由 (或是適用法條) 的詞。這一個方法，則是幫助我們簡化個別案例，希望能夠降低資料庫的大小和加速資料的比對所需花費的時間。

✓ 研究結論

我們完成了可以將一人一罪案件類型的簡易法庭案件的分類系統。依照系統的設計參數的不同，我們的系統的表現也不同，但確定的是我們可以利用程式找出判例的重要資訊，這一些重要資訊在案由判斷上可以達到將近八成的正確率。(請參考本報告所附的論文裡許多相關的細節。)

除了試驗的正確率不錯之外，比較重要的結論其實是我們認為，以機器完全自主的產生案例資料庫雖然不是一個可行的方案，不過，我們可以由機器產生案例資料檔案，再由法學專家過濾，這樣可以提高建立案例資料庫的效率。因此，我們覺得法資訊學是一個值得繼續投資人力和時間的研究方向。

附件

學術會議論文三篇及一篇碩士論文。分別發表於：人工智慧學會主辦之第七屆人工智慧與應用研討會，國際人工智慧與法資訊學學會所主辦之第九屆人工智慧與法資訊學國際研討會 (實際上，我們的文章是該研討會歷來第一篇以中文為標的論文)，第十四屆智慧型系統與方法學研討會。後面兩篇論文均會被 EI 所收錄。

由於國科會規定繳交報告比需用 .doc 或是 .rtf 檔案；而我們的論文因為格式和國科會格式不同，不方便合併在一個 .doc 檔案中。所以下列附件僅僅附上可以下載論文的網址，爾後若網址有所異動致論文不能下載，可以和報告人 (chaolin@nccu.edu.tw) 索取論文。

1. C.-L. Liu and C.-T. Chang. Some case-refinement strategies for case-based criminal summary judgments, LNAI 2871: Proceedings of the Fourteenth International Symposium on Methodologies for Intelligent Systems (ISMIS'03), to appear. Maebashi City, Japan, 28-31 October 2003. (EI, SCI Expanded)
<http://www.cs.nccu.edu.tw/~chaolin/papers/ismis032.pdf> (六頁)
2. C.-L. Liu, C.-T. Chang and J.-H. Ho. Classification and clustering for case-based criminal summary judgments, Proceedings of the Ninth International Conference on Artificial Intelligence and Law (ICAIL'03), 252-261. Edinburgh, Scotland, UK, 24-28 June 2003. (EI)
<http://www.cs.nccu.edu.tw/~chaolin/papers/icail03.pdf> (十頁)

3. 張正宗，何君豪及劉昭麟。我國簡易刑事判決的製作輔助系統 (Decision support for criminal summary judgment)，第七屆人工智慧與應用研討會論文集 (TAAI'02)，178-183。台中，台灣，15 November 2002。(中文內容)
<http://www.cs.nccu.edu.tw/~chaolin/papers/chang02.pdf> (六頁)
4. 張正宗，電腦輔助簡易刑事判決技術之探討，國立政治大學，資訊科學系碩士論文，2003。
(碩士論文電子版可由國家圖書館取得)