



大數據與法資訊學

——機器提取裁判內容要素之實踐

■邵軒磊 臺灣師範大學東亞系助理教授
吳國清 中央警察大學資訊管理學系教授
黃詩淳 臺灣大學法律系副教授

摘要

本文以資訊技術處理63,234件法院毒品相關裁判之巨量資料，目的係實現「法院裁判書」之文本探勘。法院裁判書由於形式複雜，且篇幅較長（超過10,000字），以往認為機器判讀較為困難。此外，在既有的裁判書資料庫檢索時，以目前有限的檢索條件，往往導致系統抓取過多裁判件數，但顯示件數卻又有上限，研究者必須再設定檢索條件、縮小查詢範圍，否則看不到上限數量以外之文本。本文採取的文字探勘方式，能使機器協助閱讀，省卻研究者搜尋裁判之時間精力，不受資料庫對裁判篇幅之限制，且更直接地針對研究者之關注點分析大量裁判書，使吾人能迅速掌握裁判中的整體現象。

以本文研究對象為例，大數據解析2014年1月起至2017年7月止我國關於毒

品之相關判決發現：一、研究者能藉由機器協助大量閱讀判決，並成功處理文字探勘；二、在本文研究範圍中，毒品犯罪多為集合性或系統性犯罪，且與毒品級數相符，越高級毒品利潤越高，越可能出現組織性犯罪。三、毒品犯罪具有高再犯率，本研究使用文字探勘法結果與法務部發表結果相近，能達到平行驗證效果。

壹、研究背景

傳統「裁判研究」的主流，多為對單篇判決之評釋，或對照比較數篇判決，進行深入的法學理論探討。近年人工智慧(artificial intelligence, AI)或深度學習(deep learning)成為熱門話題，坊間亦有使用AI對做出各種應用的嘗試，如學習下棋。¹在實務界，也有結合大數據犯罪偵防的實用。²

DOI: 10.3966/207798362018050071006

關鍵詞：大數據、毒品判決、法資訊學、文字探勘、法院判決

¹ David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander

資訊工具的使用，也為法學研究帶來曙光；不過，至今為止的研究其樣本多以千件左右的法院裁判書為對象，其相關研究議題主要集中在語料庫的建立、資訊檢索、中文斷詞和機器學習等技術的改進，以及內容導引式演算法的效能比較³。近期法學相關的嘗試，是使用機器學習預測裁判結果。⁴

近年法院受理案件逐漸增加⁵，若法官欲貫徹「相同之事物為相同之處理」的原則，為求謹慎，所需閱讀裁判書的量也急速增加。即使已有裁判書資料庫⁶，能夠讓使用者輸入一定條件（例如案由、日期、關鍵字等）來搜尋所需裁判，但目前之介面所能設定之條件種類有限，導致系統抓取過多裁判件數，但能實際顯示之件數卻又有上限⁷，研究者必須再設定檢索條件、縮小查詢範圍，否則看不到上限數量以外之文本。還必須花費大量精力篩選挑掉無關

的裁判，才能進行下一步的閱讀與分析。以關鍵字搜尋為例，若在此欄位輸入「毒品」二字，則只要裁判全文中出現過「毒品」二字者，均會呈現而出，即便主要爭點與毒品無關，只不過是原告或被告之主張，還是會被系統抓出，並將全部裁判依照法院與時間序列排列，使用者須一一再點選各該裁判的連結，進入後閱讀內文，才知道此一裁判是否符合自己的需求，效率甚低。

本文嘗試使用資訊科學的文字探勘(text mining)技術，以機器一次閱讀並分析數萬篇相同主題之判決，以巨觀的視角進行裁判研究。相較於傳統裁判評釋重在個案精準的解讀與分析，文字探勘方式能處理巨量資料，節省研究者瞭解背景之時間精力，直接鎖定研究者關注之問題點。以下【表1】簡要說明本文方法與資料庫搜尋及傳統裁判評釋之差異。

Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel & Demis Hassabis, Mastering the game of Go with deep neural networks and tree search, Nature, volume 529, pages 484-489 (28 January 2016)

² 編輯部，科技緝毒！新北警與IBM合作大數據分析溯源抓「毒頭」，東森新聞網，<https://www.ettoday.net/news/20180423/1155799.htm>。

³ 例如林琬真、郭宗廷、張桐嘉、顏厥安、陳昭如、林守德，利用機器學習於中文法律文件之標記、案件分類及量刑預測，中文計算語言學期刊，17卷4期，2012年12月，52頁，分析樣本為2,113件裁判。

⁴ 黃詩淳、邵軒磊，運用機器學習預測法院裁判：法資訊學之實踐，月旦法學雜誌，270期，2017年11月，86-96頁，分析樣本為448件裁判。

⁵ 司法院，中華民國105年司法統計年報，2016年，地方法院辦理刑事案件收結情形(<http://www.judicial.gov.tw/juds/year105/09/063.pdf>)，2007年受理件數481,447件，2016年受理件數為527,386件，期間各年雖有小幅消長，但大致呈現增加狀態。

⁶ 目前唯一免費開放之裁判書查詢系統為司法院之法學資料檢索系統<http://jirs.judicial.gov.tw/Index.htm>。其他民間業者所開發之裁判書查詢系統，允許更細緻的條件設定，能節省使用者精力，快速找到所需裁判，惟須付費，且其收錄之裁判書多與司法院系統相同。

⁷ 例如上述司法院之法學資料檢索系統，只能顯示查詢結果的前500筆。

【表1】裁判書之使用方式比較

	資料庫搜尋	法資訊學文字探勘	傳統裁判評釋
主要作法	由研究者在既有資料庫，設定關鍵字或案由等條件檢索，由系統回應相關裁判。	取得相關判決後，使用文字探勘技術，對關鍵知識點做因素提取。	取得相關判決後，由研究者自身經驗與加以分析。
優點	能回收大量結果。	能大量自動標記，減少研究者對議題的探索時間。	解讀較準確，能得到較為權威性之判讀結果。
缺點	需要人工閱讀內文與校正搜尋結果，再進行後續的量化或質性分析。	目前尚未有相關介面，使用者需要較高的技術能力。	受限於研究者精力，能閱讀之裁判數量有限。

貳、研究方法

爲了取得同一類型且數量充足的裁判書，且兼顧社會對研究對象的關注度，本文選擇了「毒品相關刑事裁判」作爲分析樣本。首先，從「司法院法學資料檢索系統」和「政府資料開放平臺」所提供的毒品案件法院裁判書（pdf和json等檔案類型），自2014年1月1日起至2017年7月31日止，共63,234件裁判書作爲研究樣本（樣本處理細節後詳）。並在SPARK大數據環境下，自行開發程式，來進行文件篩選，文件內容清理、資料與編碼轉換與萃取等自動化工作，並結合文字探勘、資料探勘和統計等跨領域技術，試圖達到下列之研究目的：

一、透過SPARK大數據技術，大幅縮短巨量裁判書文件所需處理時間。

二、了解臺灣地區近幾年來各級毒品與各地方法院判決案件數。

三、透過文字探勘技術，來獲取跨境毒品案件詞語與文字分類，試圖找出毒品裁判書的關鍵詞，以推測臺灣各縣市毒品犯罪類型、特徵、毒品涉及其他集團犯罪之狀況等。

初級資料蒐集之方法，係在2017年10月15日，從「司法院法學資料檢索系統」所提供的「裁判書」查詢項目，和「政府資料開放平臺」所提供法院毒品案件裁判書，在「裁判案由」鍵入「毒品危害防制條例」所得之裁判書檔案，並限定於地方法院之裁判。由於每年裁判數量極爲龐大⁸，因此進一步使用下

⁸ 首先，在數量方面，根據《中華民國105年司法統計年報》，2016年地方法院審理的刑事第一審案件，該年終結者共181,995件，其中罪名係違反毒品危害防制條例者有38,319件，僅次於公共危險（61,687件）及妨礙交通安全（58,798件）；此外，同年地方法院審理的刑事第一審簡易判決案件，該年終結者共116,869件，其中罪名係違反毒品危害防制條例者有者有19,364件，僅次於公共危險（56,376件）。將二者加總，毒品相關案件共57,683件，占全體（298,864件）之19.30%。由此可知，毒品犯罪相關的裁判書在數量上相當豐富。此外，上述數據僅是法院裁判的件數，一件裁判的被告人數未必只有一人，若以犯罪受刑人的人數觀之，根據《中華民國法務統計年報（105年）》之監獄新入監受刑人前十大罪名統計，因違反

列方式縮小：針對2014年1月1日至2016年12月31日之裁判採用隨機取樣方式，至於2017年1月1日至2017年7月31日之裁判則採全部取樣，共得本研究之樣本數爲63,432件。

其次，本文欲探究的是毒品犯罪案件的事實特徵，因此不計入裁定，僅取判決（裁定通常是程序事項，不涉及實體判斷）。再者，爲避免同樣的事實被重複計算，同一案件只計入一次（第一審裁判），排除上訴審，且無罪判決亦不計入。排除裁判書內容重複、上訴審裁判、延長羈押、保證金、司法管轄錯誤及無罪判決後，最後取得本研究有效判決文本文數爲59,041個檔案。經本研究人員撰寫SPARK & SCALA程式自動篩選，及使用了SAS Enterprise Guide、Enterprise Miner (Text Miner)做出後續計量。

參、研究成果

以下是在本研究樣本的毒品判決中，由大數據工具所觀察到之幾個有趣之現象。

一、國內毒品案件管轄法院

使用文字探勘，從「裁判字號」的欄位，可得知管轄法院爲何，進而結合統計技術，知悉各法院的案件量。在本研究之分析對象59,041筆裁判中，以臺北市（含臺北和士林等兩個地方法院）案件數(7,218)最高，占全部的12.2%；其次爲新北的6,509件；三個法院合計13,727件，占全部的23.2%。此外，苗栗

爲六都以外案件數最多的縣市，計3,841件，占全部的6.5%。

而若以毒品級數區分：就縣市地區第一級毒品犯罪案件數統計，由大至小排名前十名：桃園（1,551件，10.53%）、南投（1,351件，9.17%）、臺中（1,301件，8.83%）、彰化（1,132件，7.68%）、新北（1,116件，7.57%）、臺北（1,085件，7.36%）、新竹（1,070件，7.26%）、基隆（917件，6.22%）、高雄（916件，6.22%）、苗栗（696件，4.72%）、嘉義（696件，4.72%）、宜蘭（690件，4.68%）。就縣市地區第二級毒品犯罪案件數統計，排名前十名：臺北（5,300件，13.25%）、新北（4,341件，10.86%）、桃園（4,317件，10.80%）、苗栗（2,971件，7.43%）、基隆（2,660件，6.65%）、臺中（2,643件，6.61%）、高雄（2,640件，6.60%）、新竹（2,597件，6.49%）、南投（2,052件，5.13%）、嘉義（1,715件，4.29%）。

若研究者以傳統的法學資料檢索系統來查詢判決，在目前的介面中，雖然能看出各法院的裁判件數，但無法在檢索結果畫面同時呈現細緻的資料，例如毒品相關判決全部爲幾件，其中幾件爲第一級毒品；而必須由使用者先查詢全部判決，得出總數後，再重新回到主畫面，在「裁判主文」檢索欄位中輸入「第一級毒品」，始能得知該法院第一級毒品相關判決數量，最後再自行計算第一級毒品占全體毒品案件之比例。

毒品危害防制條例而入監者，2016年有10,933人，占全體之31.7%，排名第一。

二、毒品使用狀況與級數程度分析

「主文」的欄位，記載了犯罪方式、毒品級數、是否為累犯及科刑結果。其中，關於犯罪方式，《毒品危害防制條例》明定了製造、運輸、販賣、轉讓、持有和施用等六種。在總案件數為59,041筆中，共計69,983人次被判刑。從下表2交叉統計分析得知：

(一) 在這四種毒品分級犯案人次多寡排名上，第二級(70.56%)>第一級(25.30%)>第三級(4.05%)>第四級(0.09%)。第一級和第二級毒品的犯罪人口占了95.86%。

(二) 在六種犯罪方式個別人統計排名上，施用(71.49%)>持有(18.83%)>販賣(5.96%)>轉讓(2.92%)>運輸(0.63%)>製造(0.17%)。

三、累犯的觀察結果與政府數據平行驗證

本研究數據顯示，在裁判書總數59,041件中，累犯案件數達27,132件，占全部案件的49.95%。此與政府所提供的累犯趨勢相符，大致合乎走向。(請

見下圖)

其他本研究之觀察到的現象，因政府所公開之相關數據尚少，在未能取得政府原始資料的現狀下，較難相互比較。不過，至少在政府機關已公告的「毒品有罪判決者累犯率」(部分)數據方面，藉由比較本文結果與政府之數據，已經達到「平行驗證」的基礎能力。

不只如此，本研究還可以提供更深一層的解析：全樣本中，第一級毒品且累犯案件計9,763件，占累犯案件之35.98%。來自第二級毒品且累犯者計22,327件，占82.29%，比起第一級毒品累犯的35.98%高出很多(由於可能單一案件涉及兩種以上毒品，因此總和比例會大於百分之百)。

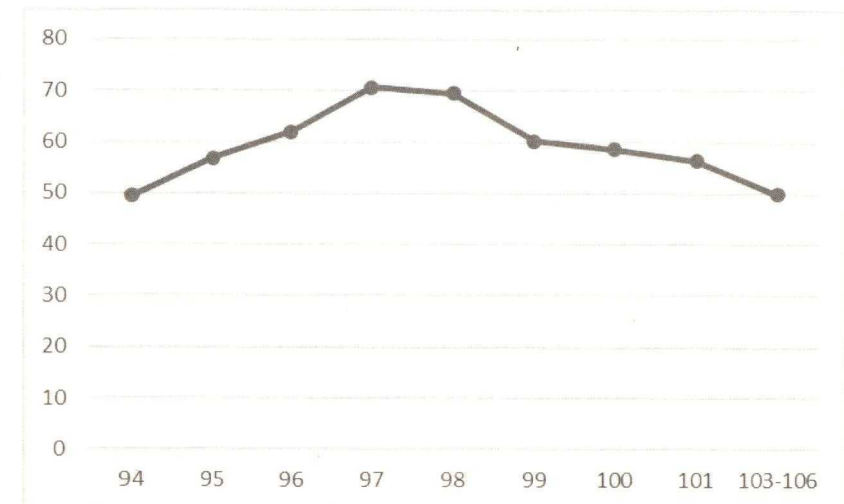
進一步分析，在「全部施用」案件41,599件中，屬於第二級毒品案件36,127件，占86.85%。在「施用且累犯」之案件中共計24,968件，屬於第二級毒品案件21,481件，占86.03%。更進一步，在「累犯施用且易科罰金」者計17,318件，又涉及第二級毒品案件計16,839件，占97.23%。上述數據表示，

【表2】毒品分級與毒品犯罪方式交叉統計分析(單位：人次)

毒品分級	製造	運輸	販賣	轉讓	持有	施用	列合計	列比例(%)
第一級毒品	25	116	978	487	2,592	13,508	17,706	25.30
第二級毒品	72	197	2,423	1,129	9,430	36,127	49,378	70.56
第三級毒品	22	113	759	426	1,136	381	2,837	4.05
第四級毒品	2	15	11	1	18	15	62	0.09
行合計	121	441	4,171	2,043	13,176	50,031	69,983+	100
行比例(%)	0.17	0.63	5.96	2.92	18.83	71.49	100	

註：由於單一案件可能涉及數人或數種毒品使用方式，因此最後合計會大於案件總數。

【圖1】毒品有罪判決者累犯率



資料來源：94-100年部分為法務部⁹，103-106年為本文大數據探勘結果。

法院對於累犯施用第二級毒品的嫌犯，大多以易科罰金為之，但第二級毒品已有相當成癮性，易科罰金很難阻止成癮者繼續吸食，累犯人數持續上升。相較於其他犯罪平均累犯率約為20~30%¹⁰，毒品的累犯比較高，是一個值得後續關注的議題。因此，綜上所述從相關判決中，本文能解析出接近標準統計的有用資料。

四、不同級毒品的集團犯罪率不同

上述文字探勘所獲之數據，能否一般常識相符？一般咸認為「毒品犯罪多

為集合性或系統性犯罪」¹¹，本研究的數據則進一步顯示，隨毒品等級的不同，組織犯罪之比例也有差異。這可能是因為利潤越高的毒品，越容易屬於共犯、共同運輸之多人參與犯罪結構，尤其是販毒集團組織犯罪。

例如下表3顯示：全樣本中涉及「共同運輸」總數220人次，其中有66人次(30%)為第一級毒品。全樣本中涉及「販毒集團」案件數計164人次，其中有84人次(51.22%)為第一級毒品。全樣本中涉及「賭博」之犯罪紀錄有392人次，其中有135人次(34.44%)為第一級

⁹ 法務部，執行裁判確定有罪者之累犯統計分析，2013年頁2圖3 (www.rjsd.moj.gov.tw/rjsdweb/common/WebListFile.ashx?list_id=932)。

¹⁰ 近8年(94年至101年)地方法院檢察署執行裁判確定有罪人數計136萬4,834人，其中累犯為35萬1,643人，占25.8%。各年累犯比率以97年之29.5%最高，94年之20.4%最低。參見法務部，同註9，1頁。

¹¹ 例如，聯合國毒品和犯罪問題辦公室，2017年世界毒品問題報告：內容提要、報告的結論和政策影響，https://www.unodc.org/wdr2017/field/WDR_Booklet1_Exsum_Chinese.pdf，便有專節討論毒品問題與組織犯罪之關聯。

【表3】毒品分級與集團犯罪方式交叉統計分析(單位：人次)

	販毒集團	共同運輸	賭博	槍砲彈藥	全樣本件數
第一級毒品集團犯罪數	84	66	135	493	17706
其他級毒品集團犯罪數	80	154	257	882	52277
總數	164	220	392	1375	69983
第一級毒品比例(%)	51.22	30	34.44	35.85	25.30
第一級毒品集團犯罪權重	3.10013	1.26536	1.55093	1.65032	

註：「第一級毒品集團犯罪權重」之計算方法為「(第一級毒品犯罪數量/其他級毒品犯罪數量) / (第一級毒品占總判決數量/其他級毒品總判決數量)」。由於單一案件可能涉及數人或數種毒品使用方式，因此最後合計會大於案件總數。

毒品案件。全樣本中涉及「槍砲彈藥刀械」者有1,375人次，其中有493人次(35.85%)為第一級毒品案件。上述第一級毒品案件的人次量所占各犯罪特徵之比例，均較第一級毒品犯案人次占全體案件人次之比例(25.30%)為高。

再更精確計算出權重後，可更明顯看出，第一級毒品在集合性犯罪(集團、共同運輸、賭博、槍砲彈藥)所占比例，相較於其他級毒品之比例為高，亦即可能有1.26~3.1倍的集團犯罪率。

伍、結論

本研究透過研究人員所開發程式語言(以Excel VBA、Spark、Scala等語言開發)與專業統計軟體，並運用文字探勘、資料探勘和統計等專業技術，從6萬餘件判決中擷取有效判決59,041件裁判書進行大數據處理與分析，得到下列之結論：

一、本研究證明資訊技術得以解析大量法院裁判文本，並萃取重要訊息，

此技術也能提供給未來其他法學領域在研究巨量資料時使用。

二、針對毒品相關判決的研究初步發現，毒品犯罪多為集合性或系統性犯罪，且與毒品級數成比例，越高級毒品愈可能是組織性犯罪。其次，毒品犯罪具有高再犯率，本研究使用文字探勘法結果，與法務部發表結果相近，實現了平行驗證。

本研究的工具不僅能用於毒品相關判決，亦能為其他主題所用，結合研究者之主體知識做出調整。相較傳統的查詢資料庫、人工篩選裁判、閱讀裁判、標記裁判之研究過程，本研究的作法更能一步到位，從巨量資料中有效率地提取對研究者有意義的資訊(information retrieval)，以便進行後續的數據分析，透過此方法所觀察到總體現象，也有助於傳統的個案研究者迅速掌握該領域的大圖像，讓研究能夠「見樹也見林」。

(本文已授權收錄於月旦知識庫及月旦系列電子雜誌www.lawdata.com.tw)



相關文獻

行政法院是否還有明天

——評釋字第758號解釋

姚葦嵐 臺灣桃園地方法院法官

摘要

我國關於行政訴訟與民事訴訟之審判，依現行法律之規定，係採司法二元訴訟制度，將性質不同之訴訟事件，基於專業及效率之考量，劃分由不同體系之法院審判。

而人民提起訴訟究應向行政法院或普通法院為之，係取決於訴訟標的之法律關係究為公法或私法。然問題乃為，此訴訟標的之法律關係究為公法或私法，應如何認定？係以該事件之原告所主張訴訟標的法律關係之法律依據為準？抑或係依該事件本質，有無涉及到公權力行使之公法行為來判斷？此於當事人涉及到國家或地方自治團體之公法人時，尤為複雜。

因國家或地方自治團體行使公權力之公法行為所生之爭執，訴訟標的之法律關係乃為公法關係，應歸行政法院；反之，因國家或地方自治團體未行使公權力之私法行為所生之爭執，訴訟標的之法律關係厥為私法關係，應由普通法院審理。於現行法仍然採行公、私法之

區分論及司法二元制之前提下，上開見解較符合行政法院及普通法院組織上專業分工之目的，本文認為應屬可採。依此，本件解釋之意旨，與釋字第540號解釋之內容有所不同，是否妥適，實容有商榷之餘地。

壹、本件釋字之事實背景及釋字內容

本件原因案件為原告葉○源以桃園市八德區公所(桃園市改制為院轄市前為桃園縣八德市公所)未經其同意即在其所有坐落於桃園市八德區之土地(下稱「系爭土地」)，鋪設柏油路面供民眾通行為由，而以桃園市政府為被告，依民法第767條第1項前段及中段規定，向臺灣桃園地方法院民事庭起訴，請求桃園市政府剷除柏油路面並返還系爭土地。惟該院以原告葉○源係依據司法院大法官釋字第400號解釋主張系爭土地尚不符公用地役權之成立要件，隱含確認無公用地役關係之請求，屬公法關係所生之爭議，應提起行政爭訟以為救濟為由，以該院104年度桃簡字第860號民

DOI: 10.3966/207798362018050071007

關鍵詞：審判權劃分、審判權衝突、審判權競合