

【專號論文】 Feature Article

DOI:10.6163/TJEAS.202012\_17(2).0003

機器學習與日本研究的對話：  
以中國社會科學院《日本學刊》為例<sup>§</sup>  
Dialogue of Machine Learning and Japan  
Studies: Analysis of Journal *Japanese Studies* of  
Institute of Japanese Studies, Chinese Academy  
of Social Sciences (JSCASS)

邵軒磊

Hsuan-lei SHAO\*

**關鍵詞：**日本研究、機器學習、自然語言處理、文字探勘、知識社群

**Keywords:** Japanese Studies, Machine Learning, Nature Language Processing, Text Mining, Knowledge Communities

---

2019年12月23日收稿，2020年3月5日修訂完成，2020年9月28日通過刊登。

<sup>§</sup> 本稿為科技部計畫「從知識系譜到知識地圖—大數據與機器學習下的『中國研究』（107-2410-H-003 -058 -MY3）」與「AI的能與不能—人文社會的脈絡、價值與效應（109-2634-F-003 -008）」之研究成果。

\* 國立臺灣師範大學東亞學系副教授

Associate Professor, Department of East Asian Studies, National Taiwan Normal University.  
ORCID: <https://orcid.org/0000-0002-7101-5272>

## 摘要

以往對某一人文知識領域的研究，往往受限研究者的個人時間與體悟靈感，如何有效率的使用時間，一直是人文學者的難題。但若加入跨領域思維，在現代科技的幫助之下，吾人可以使用數位工具對某一領域做出宏觀視角，幫助學者歸納「大量知識」。具體而言，本文嘗試結合跨領域研究手法，機器學習（Machine Learning, ML）與自然語言處理（Natural Language Processing, NLP）等，來處理《日本學刊》（中國社科院日本研究所學誌）之三十七年間二三三七篇論文，並歸納《日本學刊》研究論文的「發表類型」、各類代表性關鍵字有哪些，與各年的「發表取向」有什麼變化。本文成功實做大量中文長文本的文字解析，並能抽取其中資料訊息，並將流程半自動化，可以達到輔助研究者快速大量理解訊息。本文能抽取出高產作者、高產機構、歷年發表情形以及日本研究的主要七類範疇「金融經濟、政治、管理、文學思想、文化社會、國際關係、社會經濟」。本研究期望能提供讀者對於某個知識領域的遠程宏觀視角，以此跨領域方式研究「東亞文明」，將讓研究者開啟更多可能性。

## Abstract

Since studies on *knowledge areas/ knowledge communities* are often limited by our labor time, research cost/performance shall hence be prioritized/selected. However, our vision could be enhanced by new interdisciplinary thinking. In this article, we employed the methods of Machine Learning and Natural Language Processing to analyze 2,237 research articles. Those articles were taken from The Journal Japanese Studies published by Institute of Japanese Studies, Chinese Academy of Social Sciences (JSCASS) between 1991 and 2017. Results from the data analysis provided the answers to two questions, namely (1) what fields can be clustered in Japanese Studies, accompanied with their keywords, and (2) the number flows of the fields by year. This study analyzed thousands of Chinese long texts, and captured the key information semi-automatically, in order to efficiently support readers to approach complex information in a simplified manner. This study disclosed information on highly productive writers, highly productive institutes, their annual publications, and the main topics of Japanese Studies, which are *Economics, Politics, Management Science, Literature/Ideology, Culture/Society, IR, and Social Economy*. It could also provide a macro perspective in the “East Asian Studies” by new interdisciplinary thinking.

## 壹、 研究背景

吾人應該如何從事「日本研究」？當然，需要研究者投入相當的時間精力，也涉及大量的閱讀理解與融會貫通，尤其在現代知識精細化的結果，成為專家所需的門檻較從前更加困難。那有沒有可能引入現今發展「機器學習」技術甚至人工智慧（Artificial Intelligence, AI）從事「跨領域」探索、研究甚至開展某些傳統方法比較困難的議題呢？本文試圖使用（Machine Learning, ML）與「自然語言處理」（Natural Language Processing, NLP）等數位方法，試圖挑戰「如何從大量文獻擷取觀念」。本文將研究近年《日本學刊》期刊中之二二三七篇日本研究相關論文中的主題。因此本研究之旨趣是關於東亞兩大文明「中／日」之間的跨域研究，也是「人工智慧（機器）／人文社科」的跨領域對話。

就本研究課題之重要性而言，「中日關係」一直是東北亞區域關係的重要基礎。尤其當中日都成為區域強權的時期，更容易民族認同集體焦慮（也伴隨著現實上對方的權力擴張），反映在「日本政治」的研究也就越多。因此，在現實政治上，即使是二戰後中日也是不間斷的相互交流、合作與衝突；在理念論述上，兩者同樣是亞洲（遠東）的代表，發展路徑有巨大差異。「日本」在現實政治上與理念論述上，都與「中國」形成一個強烈的參照點。這可以說是目前學說與共同知識的主流見解，也是中國學者在做日本研究時的「公約數」；不過，如果視野放的太廣，往往容易流於泛論。因為在漫長複雜的歷史與交流過程中，吾人很容易可以找出各種合作的事件或時刻來描述兩者「一衣帶水」的和善狀態，也常能找出要對方「敦促反省」的緊張時刻。

綜上所述，為了更為聚焦鎖定現今狀態，本文期望從「中國社會科學院日本研究所」的學刊《日本學刊》來研究。之所以選取此一範圍的原因，在於相對於「研究特定人物的日本觀」或是「中國大陸社會的日本觀」而言，前者較為偏向使用特定個人的片段資料，後者需要對社會進行大規模普查，兩種方式都已經有相當研究（詳請參考「文獻回顧」節）。本文採取折衷方案，鎖定「對日代表性智庫」的觀點，期望有對上述研究之補充參考之處。而且，依照現在中國大陸的政府體制「中國社科院」屬於中國

官方之日本研究重要智庫，其言論若干程度反映政府立場，因此其日本研究著作可作為兩國關係的「指標」。

討論「兩國關係」除了政治經濟議題外，近年來的研究取徑也包括跨界現象<sup>1</sup>、人員流動<sup>2</sup>、以及知識系譜<sup>3</sup>等等。因此，吾人從「中國的日本研究的研究」切入，不只是同時代的對立與補充的學說索引，更是其他觀點的變遷脈絡與社會意義之基礎。其中不僅有認識論與方法論上最基本的問題，而且對於學科規範的實際建立制度層面，擴展到這種認識如何發展並與社會互動，再到對國際政治格局對策（政策）的制約與反饋。因此本文除了學術知識層次的整理之外，也是一個跨領域的實踐研究成果。

需要進一步闡明的是，在本文中「中國之日本研究」、「中國之日本觀」、「中日關係」三個觀念的互動。在學科關懷上，筆者會將本文所屬之研究學門放在「中日關係」上。因此本文的一個終極關懷是用來「理解中日關係」。理解基礎是從「中國的日本觀」切入，相對於使用「貿易數值」、「高層互動」、「領導政策」等方法，本文選擇使用了「中國的日本觀」。隨後，論證「中國的日本觀」之實證基礎為「中國之日本研究」，相對於「問卷調查」、「名人印象」或「文化概論」（詳後節「文獻回顧」），本文使用了「學者的研究文獻」來作為分析資料。更進一步，相較於傳統大量閱讀的方法，本研究使用「數位方法」來閱讀數千篇文獻，是本文的主要旨趣。

具體而言，本文將使用《日本學刊》自一九九一至二〇一七年發表之學術論文，並用自然語言處理以及機器學習之後，能做出自動化文本分群，再由專家歸類詮釋。本文期望達到一種對日本觀點的「遠程視角」。為此，本文使用「機器學習」方法，來研究：「中國社科院日本研究所的日本觀」並回答下列三個問題有：

- 
- 1 Fred Dallmayr, *Bordering Crossing* (Lexington: Lexington Books, 1999).
  - 2 David Campbell, Michael J. Shapiro, *Moral Spaces* (Minneapolis: University of Minnesota Press, 1998).
  - 3 David Shambaugh, *China Watching* (London: Routledge, 2006). 邵軒磊：〈中國研究議程之系譜——以日本國際政治學會誌為例〉，《問題與研究》，第 51 卷第 3 期（2012 年 9 月），頁 23-54。

- (一) 誰(人/機構)在《日本學刊》上發表?
- (二) 《日本學刊》研究論文有「發表類型」? 各類代表性關鍵字有哪些?
- (三) 各年的「發表取向」有什麼變化?

## 貳、文獻回顧

以往研究「中國之日本研究」這個題目，研究者可能以「名人日本觀研究」切入。如對單一個人，如：黃遵憲（1848-1905）、戴季陶（1891-1949）等人物之日本觀，多半是基於對當時同時通曉中日兩種文化的先進，體會日本在東亞文明位置的感悟，並與中國位置比較。比如說以黃遵憲駐日期間所見所聞用詩筆寫成的《日本雜事詩》來觀察其日本觀，多數都是使用原作者對於黃遵憲的理解，並且挑選關於原作者「認為」重要的詩句而成。<sup>4</sup>這種文獻的對話對象，必須也要非常熟悉黃遵憲的詩句，同時也要接受原作者選定重要詩的立場。還有關於戴季陶的研究，則是以其著作戴季陶的《日本論》與一九一九年其撰述的〈我的日本觀〉一文做比較，原作者認為戴對於日本的觀察，其實也是在建構對於理想中國的一種「認同」(Identity)。<sup>5</sup>這樣的論文，乃是建立在對單一人物的深刻理解之上，但若不是不熟悉研究對象的讀者將很難進入論述脈絡。再者比如說，使用比較思想法研究梁啟超（1873-1929）、林獻堂（1881-1956）、戴季陶三人，分別代表民國初年的三種華人回應日本觀點，雖然立場各有差異，但都代表了「中華文化如何藉由日本路徑理解現代化」。<sup>6</sup>

另一種研究，主詞往往是「全體中國文化的泛稱」，如「古代中國人的日本觀」<sup>7</sup>等等。這類的論文旨在涵射廣泛論點，因此多屬於統整式的概

---

4 黃濤：〈論《日本雜事詩》與黃遵憲的日本觀〉，《臺灣源流》，第54、55期合併號（2011年3月），頁119-125。

5 侯智元：〈戴季陶的日本論：思想史視野的觀察〉，《思與言：人文與社會科學雜誌》，第43卷第1期（2015年3月），頁71。

6 徐興慶：〈近代中國知識份子的日本經驗——比較梁啟超、林獻堂、戴季陶的日本觀——〉，《台大日本語文研究》，第21期（2011年6月），頁179-221。

7 汪向榮：〈古代中國人的日本觀〉，《世界歷史》，1981年第5期（1981年5月），頁33-40。

述作品。不過，此種表述方式往往無法回答具體問題，比如說「中國的日本觀」具體而言是什麼？又受什麼原因影響？往往是在作者研究諸多資料之後，得到自身感觸。毋寧說在「中日關係」這個議題上，任何論述都可能找足夠充分的論證的前提下，立場甚至論述本身就成為「主體選擇」問題。換言之，中日學者互相研究彼此的同時，也是相互把對方他者化，最終反映出對自己國家的看法。更具體而言，研究「中國的日本觀」能夠用來更瞭解中國。廣義而言，研究某主體的任何觀點，皆為對某主體的研究。

第三種是屬於調查研究，這種往往需要相當巨大的研究成本。如大型民調專案《中美公眾的世界觀念調查》以問卷調查的方式，同時在中美兩國做了全國性民調：「〔……〕中國部分隨機發放邀請 33000 份，回應數為 3551 人，成功樣本為 2000。美國部分發放邀請 35000 份，有 7500 人回應，成功樣本為 2200。」<sup>8</sup>在得到問卷資料之前，常必須投入大量成本。這就是大規模之社科方法，往往需要機構等級以上的研究資源，構成了研究門檻。

上述的「日本研究」文獻可分為三類，其一是對少數名人的思想研究；其二是較廣義的文化研究；其三是社科的問卷研究。三者都有其獨到之處，但也各有若干限制。本文在綜合以上的路徑與討論方式，筆者認為可能提出「跨領域／機器學習」的研究法，研究「中國社科院日本研究所的日本研究」，從而提出較上述研究更有「普遍性」「重複性」能力的觀點，同時嘗試符合本刊「跨領域新思維」的旨趣。與本文類似討論文本探勘論文的研究，比如國外之政治學者提出「Text as Data」的觀念，將「政治文本」視為某種可能被量化解析之資料。<sup>9</sup>亦有實際應用至大量文本的解讀，如研究了人民日報報導中的人權主題，<sup>10</sup>或是與本文旨趣類似，研究臺灣學刊如

8 張昆、王宇婷：〈中美兩國公眾心中目的日本國家形象——基於 2016 年中美兩國的全國性民意調查〉，《新聞大學》，2018 年第 5 期（2018 年 5 月），頁 134。

9 Justin Grimmer, Brandon Stewart, "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts," *Political Analysis*, Vol. 21, No. 3 (Jan., 2013), pp. 1-31.

10 Titus Chiche Chen, Chiahao Hsu, "Double-Speaking Human Rights: Analyzing Human Rights Conception in Chinese Politics," *Journal of Contemporary China*, Vol. 27, No. 112 (Feb., 2018), pp. 1-15.

何研究「中國大陸」。<sup>11</sup>在傳播學中，也有學者注意到演算法對於未來訊息傳播的影響，<sup>12</sup>上述都是其他研究領域上之跨領域研究突破。

## 參、 研究設計

### 一、研究資料

本文選取：中國社科院日本研究所（以下簡稱日本所）的期刊論文作為「語料庫」。日本研究所成立於一九八一年五月，是中國社科院下屬的專門從事當代日本問題研究的學術機構，是中國大陸對於日本研究的代表性智庫，其概況為：

在編制上，日本研究所編制為 58 人，其中研究員為 43 人，科研職輔助為 8 人，行政管理職為 7 人。研究員除了本薪之外，也根據國家經濟社會發展和當代哲學社會科學學科發展的需要，承擔國家社科基金專案、中國社科院創新工程的重大專案、研究所創新工程項目、國情調查專案、各類委託交辦專案等。日本所由日本政治、日本外交、日本經濟、日本社會、日本文化等五個研究室和《日本學刊》編輯部、網路資料室、綜合辦公室構成。日本研究所設置日本政治研究中心、中日關係研究中心、中日經濟研究中心、中日社會文化研究中心等四個非實體跨所研究機構，代管中華日本學會和全國日本經濟學會等兩個全國性日本研究學術團體，也能招收和培養當代日本研究的碩士生和博士生，設立當代日本研究博士後流動站。<sup>13</sup>

- 
- 11 邵軒磊、曾元顯：〈文字探勘技術輔助主題分析——以「中國大陸研究」期刊為例〉，《問題與研究》，第 57 卷第 1 期（2018 年 3 月），頁 29-62。
  - 12 蕭維傑、王維菁：〈動態消息演算法、能見度、與新聞經營：以 Facebook 平台為例〉，《資訊社會研究》，34 期（2018 年 1 月），頁 63-104。
  - 13 中國社科院日本研究所：〈本所簡介〉，<http://ijs.cssn.cn/bsgk/bsjj/>，檢索日期：2020 年 4 月 3 日。

而其刊物《日本學刊》既為其組織學刊，也是「中華日本學會」之會誌，可以大致上代表中國大陸日本研究學界之研究觀點，而學術資料可以公開自期刊論文資料庫取得，是客觀存在的資料。本文使用《日本學刊》自一九九一至二〇一七年之發表文章，原始論文共有三二二二篇，但經過清洗調整（如僅有篇名而無公開內文、僅為記事或報導而非論文）後，可用資料有二二三七篇學術文獻。

## 二、機器學習

在研究方法上，本文使用機器學習中之非監督式文本分類。<sup>14</sup>「文本相似」是一個抽象觀念，研究者可以藉由自身需求而定義並賦予合理解釋。有的研究會使用「字詞數長短」、「使用某些特殊字（如感嘆詞）」、「相似詞詞對數」等等來解釋。而本文使用「潛在狄立克雷分配」（Latent Dirichlet Allocation, LDA）演算法，使用貝式分配（Bayesian theory）計算某些詞組分配能更降低整體語料的「亂度」（Entropy）。本文就將這種分配認為是「相似標準」從而歸類相似的文件主題。就本次所研究樣本特性（長文本、數量為數千筆），使用 LDA 得到不錯的效果。雖然隨著樣本變化，「分類」結果也會發生微調，因此本文並不宣稱是一個「永遠的解答」。相對於此，本文之主要問題意識在於使用「跨領域技術與思維」，能以此自動化（或半自動化）分群，提供領域研究者觀察學科的初步觀念，並且探索中文文本，並提供未來探索人工智慧的前製基礎知識作業。

LDA 非監督式學習（Unsupervised Learning）之特性為其專擅之處。<sup>15</sup>至今，也已被廣泛延伸到其他領域，如圖像辨識等等。其要訣在於使用貝式定理，找出最有可能的組合主題，因此也能用於字詞組合。<sup>16</sup> LDA 主要的核心概念為，自然直觀的現實世界中，吾人在閱讀時能將文章視為「字詞→文章」的原子化組成，也可以使用一種方式為「字詞→主題→文章」

14 非監督式文本分類意指「沒有標準答案」的文本分類，又稱分群（clustering），演算法透過尋找資料的特徵，自己進行分類。因此我們能找到誰比誰更相似，在本文中將相似的一群文章解釋為「擁有共同主題」。

15 David M. Blei, John D. Lafferty, "A Correlated Topic Model of Science," *Annals of Applied Statistics*, Vol. 1, No. 1 (Jan., 2007), pp. 17-35.

16 David M. Blei, Andrew Y. Ng, Michael I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, Vol. 3, No. 4-5 (Apr., 2003), pp. 993-1022.

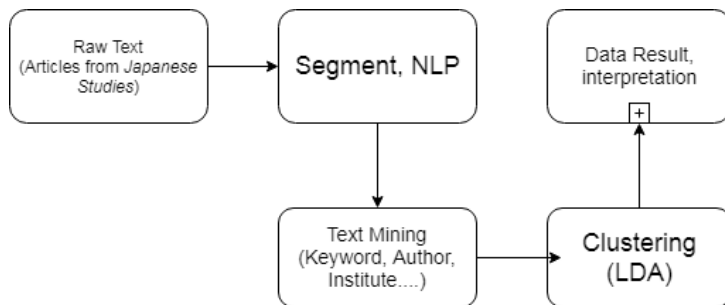


的三層模型。這樣的模型可以使用某種機率數學式來表示。在特定有限文章（語料庫）中，主題數量、字詞範圍都能確定，因此有一種「分配」能使其語料之交叉亂度（Cross-entropy）總和最小，<sup>17</sup>亦即某形式分佈的主題最有「解釋力」，比如說後表所示（金融、邦交、經濟學、哲學／憲法、行政、戰爭、民族主義）等等的分群，是經過某種計算後，最有效減少前述交叉亂度總和的一種分類，遂由此得以成立「最佳主題分群」。

### 三、研究流程

由於對於以「字詞」為主的各式演算，由於中文字詞本身之歧義性，研究者也常會使用本身主體知識進行理解。因此，使用數位工具主要是提供方便使用者，能夠更快也更多掌握文本資訊，但同時要容許讀者各以其認知角度和需求來閱讀和解讀文獻。雙方皆會對某一主題，從其自身專業知識進行深度描述或詮釋，意即文本與讀者之「互為主體」。

總體而言，本文之研究流程為「蒐集文本→輸入製作語料庫→分詞→文字探勘關鍵訊息（作者、機構）→機器學習分群→詮釋結果」。茲將流程圖示如下：



【圖一】本文研究架構<sup>18</sup>

17 在資訊科學中，增加資訊大多會增加模型的預測率。比如說「早上天上有烏雲」就會增加「下午下雨」的機率，這樣我們就說這降低了模型的亂度。但是「我早上吃麵」對「下午下雨」幾乎不會有影響。我們就說「我早上吃麵」無法降低模型亂度。

18 資料來源：由作者自行繪製。

## 肆、研究成果與討論

本次研究蒐集了一九九一至二〇一七年，共二二三七篇《日本學刊》的文章，期望回答：「中國社科院日本研究所的日本觀」並回答下列三個問題：

- (一) 誰（人／機構）在《日本學刊》上發表？
- (二) 《日本學刊》研究論文有「發表類型」？各類代表性關鍵字有哪些？
- (三) 各年的「發表取向」有什麼變化？

以下分為數節呈現研究成果：

### 一、誰／機構能在《日本學刊》上發表？

依照論文之資料之作者欄位，使用文字探勘去除雜訊，並整理出有效作者姓名，並依照寫作「篇數」，整理高產作者如下表：

【表一】《日本學刊》發表作者篇數頻次表<sup>19</sup>

rank	author	Freq	rank	author	Freq	rank	author	Freq
1	編輯部	165	21	張舒英	15	41	許曉光	9
2	林昶	79	22	李璿夏	14	42	陳建安	9
3	葉琳	37	23	張義素	14	43	裴桂芬	9
4	蔣立峰	32	24	劉世龍	14	44	劉瑞	9
5	江瑞平	30	25	劉德有	14	45	何方	8
6	馮昭奎	30	26	王偉	13	46	吳寄南	8
7	王新生	24	27	李卓	13	47	李秀石	8
8	劉江永	23	28	尚會鵬	13	48	林曉光	8
9	金熙德	22	29	高增傑	13	49	武心波	8
10	高洪	22	30	范作申	11	50	金仁淑	8
11	徐梅	21	31	孫承	11			

19 資料來源：由作者自行繪製。

12	崔世廣	21	32	張伯玉	11		
13	張季風	21	33	張建立	11		
14	劉昌黎	20	34	莽景石	11		
15	韓鐵英	20	35	廉德瑰	11		
16	呂耀東	19	36	葉渭渠	11		
17	胡澎	19	37	王中田	10		
18	武寅	18	38	李文	10		
19	吳懷中	17	39	韓永順	10		
20	徐萬勝	17	40	孫執中	9		

「編輯部」是最多的，代表全體機構立場寫作的文獻，因為很多文獻並無具名，或是以全所名義發表的，在本文統一使用「編輯部」。就具名的作者而言，前十名是林昶、葉琳、蔣立峰、江瑞平、馮昭奎、王新生、劉江永、金熙德、高洪。主要是所內研究員：如林昶，曾任中國社會科學院日本研究所所長秘書、現為《日本學刊》編輯部主任。蔣立峰、馮昭奎、金熙德、高洪是日本研究所學者，也常常在媒體上發表見解。不過也有外校學者，如葉琳、江瑞平。

而更進一步，將每個作者投稿所屬機構數量表列如下：

**【表二】《日本學刊》發表機構篇數頻次表<sup>20</sup>**

rank	institute	Freq	rank	institute	Freq	rank	institute	Freq
1	日本學刊編輯部	620	21	上海外國語大學	9	41	青島農業大學	5
2	中國社會科學院	539	22	北京師範大學	9	42	首都師範大學	5
3	其他	160	23	四川師範大學	9	43	雲南大學	5
4	南開大學	89	24	上海交通大學	8	44	對外經濟貿易大學	5
5	復旦大學	77	25	上海財經大學	7	45	遼寧師範大學	5

20 資料來源：由作者自行繪製。46名機構以後為4篇，有超過10個機構，如曲阜師範大學、延邊大學、杭州大學等，因此採用並列形式，不計名次。

6	北京大學	75	26	同濟大學	7	〈45	中國現代國際關係研究院等（並列）	4
7	吉林大學	68	27	武漢大學	7			
8	河北大學	59	28	南京大學	7			
9	遼寧大學	57	29	廣東外語外貿大學	7			
10	東北師範大學	51	30	上海師範大學	6			
11	天津社會科學院	32	31	中山大學	6			
12	東北財經大學	32	32	國際關係學院	6			
13	外交學院	31	33	朝鮮社會科學院	6			
14	上海國際問題研究院	26	34	廈門大學	6			
15	中國人民大學	21	35	暨南大學	6			
16	清華大學	19	36	大連外國語學院	5			
17	中國政法大學	13	37	山東大學	5			
18	解放軍外國語學院	13	38	中國傳媒大學	5			
19	上海社會科學院	11	39	東京大學	5			
20	華東師範大學	11	40	青島大學	5			

從機構而言，除了日本研究所與中國社科院外，可以看出南開、復旦、北大、吉大、河北大學等，都是十分活躍的日本研究機構。

## 二、所有論文分詞、研究取向分群與其關鍵字

吾人先將所有論文之「題目、關鍵字、摘要」作為一整個語料（Corpus），隨後將語料進行分詞，如下表所演示：

【表三】分詞範例表<sup>21</sup>

原文範例	分詞範例
目前中日兩國進行的戰略博弈有三點值得關注：一是戰略博弈的歷史定位。在兩千多年中日關係史上，世紀中後期第一次出現雙方實力對比關係逆轉和雙邊格局重構。	目前、中日、兩國、進行、的、戰略、博弈、有、三點、值得、關注、一是、戰略、博弈、的、歷史、定位、在、兩千多年、中日、關係史、上、世紀、中後期第一次、出現、雙方、實力、對比、關係、逆轉、和、雙邊、格局、重構

吾人可以在上述範例的分詞中，找到重要的常用詞如「中日、戰略、博弈」，大致上可以了解到越常出現的詞，多半越是本文的「特徵觀念」。以肉眼可以看出，演示範例與我們閱讀的經驗大致相同，本段文本主要說明中日戰略。因此，本文可以更進一步將二二三七篇文獻斷詞並統計出詞頻表，可以看出「所有」語料的特徵詞，如下所示：

【表四】《日本學刊》所有詞頻表<sup>22</sup>

rank	Var1	Freq	rank	Var1	Freq	rank	Var1	Freq	rank	Var1	Freq
1	經濟	2315	26	結構	252	51	體系	167	76	國民	118
2	政治	1100	27	中心	247	52	市場	165	77	形勢	115
3	文化	1064	28	因素	245	53	基礎	165	78	民族主義	114
4	關係	936	29	產業	226	54	行政	162	79	民間	112
5	國際	928	30	時代	223	55	內閣	161	80	背景	111
6	問題	903	31	方面	211	56	方式	158	81	內容	110
7	戰略	858	32	地區	208	57	日元	155	82	泡沫經濟	109
8	研究所	745	33	理論	207	58	階段	154	83	經歷	107
9	社會	726	34	金融	202	59	戰爭	149	84	大學	106
10	政策	704	35	傳統	200	60	技術	148	85	軍國主義	106
11	國家	669	36	學者	198	61	軍事	147	86	理念	106
12	企業	643	37	憲法	197	62	主義	146	87	決策	105

21 資料來源：由作者自行繪製。

22 資料來源：由作者自行繪製。

13	歷史	577	38	環境	194	63	基本	144	88	精神	105
14	外交	556	39	模式	191	64	目標	142	89	水準	101
15	學刊	472	40	會議	189	65	民族	141	90	課題	98
16	制度	446	41	文學	186	66	農業	139	91	代表	97
17	學術	409	42	特點	185	67	機構	137	92	條約	97
18	世紀	383	43	原因	181	68	格局	135	93	經驗	97
19	世界	344	44	意識	181	69	地位	129	94	財政	96
20	體制	296	45	邦交	178	70	政黨	129	95	政權	95
21	過程	292	46	產生	175	71	區域	129	96	公司	91
22	政府	286	47	學會	175	72	意義	127	97	哲學	91
23	思想	284	48	利益	174	73	領域	127	98	轉型	90
24	時期	283	49	特徵	174	74	機制	127	99	交流	89
25	研討會	263	50	經濟學	169	75	進程	121	100	全國	89

從詞頻中看來，「經濟」是遙遙領先其他詞頻，共有二三一五個，足見主要上開刊物《日本研究》的主要關心就是經濟。而「政治」、「文化」、「國際關係」、「社會」等各領域代表詞，也都在前十名內。原因在於，這些大致屬於「學科性描述用詞」，其範疇較大。五十至一百名的詞彙中就有比較細項的詞，如「內閣、日圓、政黨」等，就是範疇較小的詞。「泡沫經濟」、「軍國主義」，是中國大陸學者研究日本的重要關鍵詞，比如說「軍國主義」，可以由類似的用詞看出作者群之研究取向。

「詞頻」能顯示所有的傾向，因此本研究能更進一步可以使用機器學習分群，將這些文件，依照文中出現之詞彙聚類出類似的組別，可以觀察更進一步的細項。在本文中先自動化分為七組，再依照人眼識別關鍵字，命名為：「金融經濟、政治、管理、文學思想、文化社會、國際關係、社會經濟」等七組。如下表所示：

【表五】《日本學刊》分群詞頻表<sup>23</sup>

Topic 1 金融經濟	Topic 2 政治	Topic 3 管理	Topic 4 文學思想	Topic 5 文化社會	Topic 6 國際關係	Topic 7 社會經濟
金融	憲法	企業	文學	文化	產業	日元
邦交	行政	技術	農業	意識	釣魚島	政黨
經濟學	戰爭	民間	資料	文明	對華	國際化
哲學	主義	公司	江戶	宗教	台灣	財政
全國	民族主義	科技	藍皮書	朝鮮	製造業	民主黨
銀行	軍國主義	工業	社會科學	婦女	安保	家庭
金融危機	靖國神社	經濟體制	小說	神道	海洋	資本主義
貨幣	小泉	價格	儒學	女性	產業結構	共同體
東京	思潮	立國	身份	佛教	能源	地方
人民	天皇制	日本式	農村	國民性	規制	政局
城市	天皇	法人	德川	藝術	地緣	物價
國際交流	右翼	市場經濟	武士	甲午戰爭	聯合國	匯率
方針	教科書	外國	幕府	形象	友好條約	人口
馬克思主義	保守主義	產業政策	輿論	帝國主義	夥伴關係	議員
東北亞	日本國	資產	差距	日語	沖繩	國債
貨幣政策	感情	範式	作品	思維	威脅論	政界
文化交流	策略	職能	投票	特性	主席	迴圈
歐洲	太郎	歐美	品質	福澤諭吉	競爭力	國會
日本銀行	倫理	科研	農產品	交流史	雙邊	景氣
貸款	議會	股票	工程	性格	福田	政治體制
存款	國家主義	汽車	韓國	史學	琉球	地震
泡沫	家族	網路	天津	語言	空心化	勞動力
學術研究	信任	序列	規範	方法論	朝鮮半島	土地
利率	史觀	科學	大陸	關係史	中曾根康弘	居民
區域合作	侵略擴張	收支	獎評	列島	政經	社會保障
學派	歷史系	員工	朱子學	規律	石油	鳩山
人士	日本首相	經驗教訓	武士道	加藤	和約	稅制

23 資料來源：由作者自行繪製。

面向	事變	法制	文學史	風土	國際法	債務
債權	權利	經貿	辦法	神話	日方	中央
日本學	三國	關係史	學術論文	印度	改革開放	石油危機

選擇分成「七」組，並不是只能分為七組，而是參考類似論文分類。在呈現上，筆者也比較了三至九種分類，但以本文選擇七組 LDA 分類的多樣性與解釋力似乎最佳。目前非監督式分群雖有若干數學指標能協助研究者之分群數量，但細節解釋仍須執筆者之經驗與專業判斷，需要研究者之主體知識詮釋數學模型結果。

### 三、所有論文分詞、研究取向分群與其關鍵字

上述兩表（分類與機構統計）結合，就能看出各個機構的發表取向，如下表所示。其中橫列項為上一小節「機器學習」做出之半自動化分群，而縱向項為各機構（原始資料經由文字探勘而呈現）。這可以幫助吾人瞭解每個機構在每個議題的發表能量，以及所佔比重。由於篇幅，在本表中取用的是前二十名機構，二十一名之後則匯聚為「其他」項呈現：

**【表六】《日本學刊》發表機構分群數量表<sup>24</sup>**

	金融 經濟	政治	管理	文學 思想	文化 社會	國際 關係	社會 經濟	總計
日本學刊編輯部	166	100	54	126	77	61	36	620
中國社會科學院	86	112	58	51	66	87	79	539
其他	44	25	21	18	15	25	12	160
南開大學	17	17	10	10	17	6	12	89
復旦大學	13	9	13	7	11	17	7	77
北京大學	8	21	5	8	20	7	6	75
吉林大學	12	6	16	9	3	14	8	68
河北大學	9	3	21	6	0	10	10	59
遼寧大學	18	2	18	7	4	4	4	57
東北師範大學	4	10	5	13	9	5	5	51

24 資料來源：由作者自行繪製。



天津社會科學院	1	6	9	8	1	3	4	32
東北財經大學	4	2	11	5	0	4	6	32
外交學院	7	6	5	0	2	6	5	31
上海國際問題研究院	3	3	2	3	2	11	2	26
中國人民大學	5	4	5	2	2	1	2	21
清華大學	2	2	0	0	3	10	2	19
中國政法大學	0	1	2	0	0	5	5	13
解放軍外國語學院	2	3	0	2	0	0	6	13
上海社會科學院	1	3	0	0	2	4	1	11

因為排名篇數首尾（top 20）差異甚大，所以亦可使用橫向百分比，來呈現各機構在日本研究各分項的發表比例：

**【表七】《日本學刊》發表機構分群百分比表<sup>25</sup>**

	金融 經濟	政治	管理	文學 思想	文化 社會	國際 關係	社會 經濟	總篇 數
日本學刊編輯部	26.77%	16.13%	8.71%	20.32%	12.42%	9.84%	5.81%	620
中國社會科學院	15.96%	20.78%	10.76%	9.46%	12.24%	16.14%	14.66%	539
其他	27.50%	15.63%	13.13%	11.25%	9.38%	15.63%	7.50%	160
南開大學	19.10%	19.10%	11.24%	11.24%	19.10%	6.74%	13.48%	89
復旦大學	16.88%	11.69%	16.88%	9.09%	14.29%	22.08%	9.09%	77
北京大學	10.67%	28.00%	6.67%	10.67%	26.67%	9.33%	8.00%	75
吉林大學	17.65%	8.82%	23.53%	13.24%	4.41%	20.59%	11.76%	68
河北大學	15.25%	5.08%	35.59%	10.17%	0.00%	16.95%	16.95%	59
遼寧大學	31.58%	3.51%	31.58%	12.28%	7.02%	7.02%	7.02%	57
東北師範大學	7.84%	19.61%	9.80%	25.49%	17.65%	9.80%	9.80%	51
天津社會科學院	3.13%	18.75%	28.13%	25.00%	3.13%	9.38%	12.50%	32
東北財經大學	12.50%	6.25%	34.38%	15.63%	0.00%	12.50%	18.75%	32
外交學院	22.58%	19.35%	16.13%	0.00%	6.45%	19.35%	16.13%	31

25 資料來源：由作者自行繪製。

上海國際問題研究院	11.54%	11.54%	7.69%	11.54%	7.69%	42.31%	7.69%	26
中國人民大學	23.81%	19.05%	23.81%	9.52%	9.52%	4.76%	9.52%	21
清華大學	10.53%	10.53%	0.00%	0.00%	15.79%	52.63%	10.53%	19
中國政法大學	0.00%	7.69%	15.38%	0.00%	0.00%	38.46%	38.46%	13
解放軍外國語學院	15.38%	23.08%	0.00%	15.38%	0.00%	0.00%	46.15%	13
上海社會科學院	9.09%	27.27%	0.00%	0.00%	18.18%	36.36%	9.09%	11

解讀時可以使用「日本學刊編輯部」為基準，前三名為為金融經濟（26.77%）、文學思想（20.32%）、政治（16.13%）。這原則上可以列為日本研究的三種主要平均取向。來自「中國社會科學院」，則較諸「日本研究編輯部」的取向有相當不同，如「政治、國際關係、社會經濟」研究數量百分比顯然較多，可以推估這是「《日本研究》編輯部」與「中國社科院」其他個人研究者的分工不同。其次，可以做出每個取向的刊出種類排名如下表，如政治類以「北京大學、上海社科院」最多；國際關係以「清華大學、上海國際問題研究院」為多：

**【表八】《日本學刊》發表分群機構排名表<sup>26</sup>**

Rank	金融經濟	政治	管理	文學思想	文化社會	國際關係	社會經濟
1	遼寧大學	北京大學	河北大學	東北師範大學	北京大學	清華大學	解放軍外國語學院
2	其他	上海社會科學院	東北財經大學	天津社會科學院	南開大學	上海國際問題研究院	中國政法大學
3	日本學刊編輯部	解放軍外國語學院	遼寧大學	日本學刊編輯部	上海社會科學院	中國政法大學	東北財經大學
4	中國人民大學	中國社會科學院	天津社會科學院	東北財經大學	東北師範大學	上海社會科學院	河北大學
5	外交學院	東北師範大學	中國人民大學	解放軍外國語學院	清華大學	復旦大學	外交學院

不過，本表可能因為母數過少，如上海社科院全部僅有十一篇，則很

26 資料來源：由作者自行繪製。

容易因為一些某些類別過高或低。所以本表排名部分僅供讀者參考，在數量較大時方可做更深解讀。

#### 四、各年的「發表取向」有什麼變化？

在前述各分類的基礎之上，若配合各年度論文可能尋找出各分類歷年發展趨勢。如下表所示：

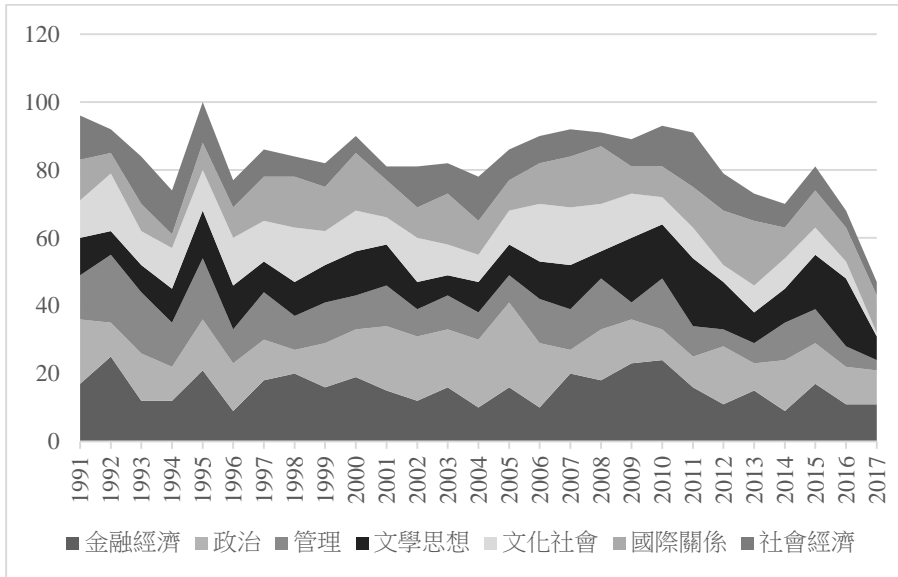
**【表九】《日本學刊》歷年分群數量表<sup>27</sup>**

	金融 經濟	政治	管理	文學 思想	文化 社會	國際 關係	社會 經濟	總和
1991	17	19	13	11	11	12	13	96
1992	25	10	20	7	17	6	7	92
1993	12	14	18	8	10	8	14	84
1994	12	10	13	10	12	4	13	74
1995	21	15	18	14	12	8	12	100
1996	9	14	10	13	14	9	8	77
1997	18	12	14	9	12	13	8	86
1998	20	7	10	10	16	15	6	84
1999	16	13	12	11	10	13	7	82
2000	19	14	10	13	12	17	5	90
2001	15	19	12	12	8	11	4	81
2002	12	19	8	8	13	9	12	81
2003	16	17	10	6	9	15	9	82
2004	10	20	8	9	8	10	13	78
2005	16	25	8	9	10	9	9	86
2006	10	19	13	11	17	12	8	90
2007	20	7	12	13	17	15	8	92
2008	18	15	15	8	14	17	4	91
2009	23	13	5	19	13	8	8	89
2010	24	9	15	16	8	9	12	93

27 資料來源：由作者自行繪製。

2011	16	9	9	20	9	12	16	91
2012	11	17	5	14	5	16	11	79
2013	15	8	6	9	8	19	8	73
2014	9	15	11	10	9	9	7	70
2015	17	12	10	16	8	11	7	81
2016	11	11	6	20	5	10	5	68
2017	11	10	3	7	1	11	4	47
總和	423	373	294	313	288	308	238	2237

以及使用疊加圖可視化，如下圖所示：



【圖二】《日本學刊》歷年分群數量表（篇數）<sup>28</sup>

由圖中可以看出，各分類基本上平均，表示《日本學刊》出刊穩定，其中高低，其中每期或有增減一兩篇，或因為資料庫緣故而僅有篇名而無摘要或關鍵字，造成差異。

總體而言，「金融經濟」為最大宗，尤其在九〇年代與二〇〇七至二〇一〇年日本經濟是日本研究中之重要議題，其關鍵字有：「金融、邦交、

28 資料來源：由作者自行繪製。

經濟學、銀行、金融危機、貨幣、人民（幣）、城市、國際交流、貨幣政策」等等。與金融經濟有關的是「管理」，其關鍵字是「企業、技術、民間、公司、科技、工業、經濟體制、價格、市場經濟」。這兩者有若干競合關係，在九〇年代之中國大陸十分風行「日本管理」方式，諸如日本民間企業文化、市場經濟法則、甚至「經營之神」松下幸之助或是儒家資本主義等等，都是中國大陸「改革開放」政策學習的對象。也有若干民間經貿往來因素，反映了初期中日經濟關係由日本民間企業帶頭的風潮。不過，「管理」類研究隨著日本經濟衰弱，中國大陸經濟超越日本後，更多轉為討論日本為何經濟不景氣，數量上也逐漸變少。「社會經濟」類研究關鍵字有：「日元、政黨、財政、民主黨、家庭、資本主義、共同體、地方、政局、物價、匯率、人口」等，若干程度比較接近社會學方面的現狀調查，與「管理」類的狀況有其類似之處。

在政治方面，中期（1990年代後半至2000年代前半）較為繁盛，若干反映了九〇年代開始的「日本熱」，至今逐漸成為另一個主流。其關鍵字有：憲法、行政、戰爭、民族主義、軍國主義、靖國神社、小泉、思潮、天皇制、天皇、右翼、教科書、保守主義、史觀、侵略擴張」等等。讀者可以從上述的關鍵字裡看出中國大陸對日本的關心議題。在這段期間中，尤其是對於小泉純一郎首相（1942-，2001年4月-2006年9月任職首相）的各種研究，可說是當時對日本政治研究的核心議題。可以一起討論的是從「國際關係」的角度來研究日本，其關鍵字有（釣魚島、對華、台灣、安保、海洋、能源、地緣、聯合國、友好條約、夥伴關係、沖繩、威脅論），則是逐漸增多，從初期不甚明顯到現今十分蓬勃。「政治」與「國際關係」的不同在於，國際關係更多涉及國際結構、國際組織、第三方國家與領土糾紛，這當然與中國大陸自身逐漸在國際上有發言權相關，同時也反映出中日政治關係逐漸國際議題化。「國際關係」在二〇〇八年、二〇一三年有明顯的高峰，則是因應具體北京奧運與釣魚島事件等中日糾紛。

兩個文史哲系列，「文學思想」關鍵字：「文學、江戶、小說、儒學、身份、農村、德川、武士、幕府、輿論、朱子學、武士道」等等。「文化社會」關鍵字：「文化、意識、文明、宗教、婦女、神道、女性、佛教、

國民性、藝術、甲午戰爭、形象、帝國主義、福澤諭吉、交流史」等等。也與「政治／國際關係」或「金融經濟／管理」相像，有若干競合關係。前期兩者不相上下，但中期「文化社會」略多於「文學思想」，現今反而大多是「文學思想」而較少「文化社會」。這與二〇〇〇年代後半的幕府熱、江戶時期的研究可能有若干關係。

吾人亦可將其百分率列出，更易比較：

**【表十】《日本學刊》歷年分群百分比表<sup>29</sup>**

	金融經濟	政治	管理	文學思想	文化社會	國際關係	社會經濟
1991	17.71%	19.79%	13.54%	11.46%	11.46%	12.50%	13.54%
1992	27.17%	10.87%	21.74%	7.61%	18.48%	6.52%	7.61%
1993	14.29%	16.67%	21.43%	9.52%	11.90%	9.52%	16.67%
1994	16.22%	13.51%	17.57%	13.51%	16.22%	5.41%	17.57%
1995	21.00%	15.00%	18.00%	14.00%	12.00%	8.00%	12.00%
1996	11.69%	18.18%	12.99%	16.88%	18.18%	11.69%	10.39%
1997	20.93%	13.95%	16.28%	10.47%	13.95%	15.12%	9.30%
1998	23.81%	8.33%	11.90%	11.90%	19.05%	17.86%	7.14%
1999	19.51%	15.85%	14.63%	13.41%	12.20%	15.85%	8.54%
2000	21.11%	15.56%	11.11%	14.44%	13.33%	18.89%	5.56%
2001	18.52%	23.46%	14.81%	14.81%	9.88%	13.58%	4.94%
2002	14.81%	23.46%	9.88%	9.88%	16.05%	11.11%	14.81%
2003	19.51%	20.73%	12.20%	7.32%	10.98%	18.29%	10.98%
2004	12.82%	25.64%	10.26%	11.54%	10.26%	12.82%	16.67%
2005	18.60%	29.07%	9.30%	10.47%	11.63%	10.47%	10.47%
2006	11.11%	21.11%	14.44%	12.22%	18.89%	13.33%	8.89%
2007	21.74%	7.61%	13.04%	14.13%	18.48%	16.30%	8.70%
2008	19.78%	16.48%	16.48%	8.79%	15.38%	18.68%	4.40%
2009	25.84%	14.61%	5.62%	21.35%	14.61%	8.99%	8.99%
2010	25.81%	9.68%	16.13%	17.20%	8.60%	9.68%	12.90%
2011	17.58%	9.89%	9.89%	21.98%	9.89%	13.19%	17.58%
2012	13.92%	21.52%	6.33%	17.72%	6.33%	20.25%	13.92%

29 資料來源：由作者自行繪製。

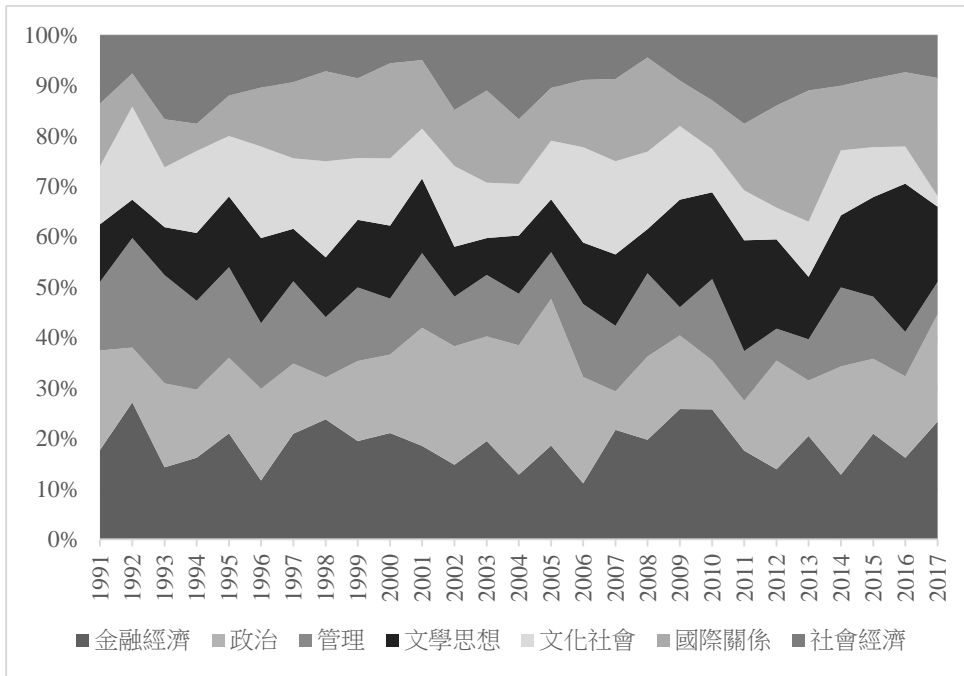
2013	20.55%	10.96%	8.22%	12.33%	10.96%	26.03%	10.96%
2014	12.86%	21.43%	15.71%	14.29%	12.86%	12.86%	10.00%
2015	20.99%	14.81%	12.35%	19.75%	9.88%	13.58%	8.64%
2016	16.18%	16.18%	8.82%	29.41%	7.35%	14.71%	7.35%
2017	23.40%	21.28%	6.38%	14.89%	2.13%	23.40%	8.51%
總平均	18.91%	16.67%	13.14%	13.99%	12.87%	13.77%	10.64%

總體而言，讀者可以藉由「機器學習」看出大量文獻所帶來的「趨勢」。在一九九〇年代，中國大陸的研究者主要還是從「經濟、商管」等等實用觀點來研究日本。有相當濃厚的「以日為師」的傾向（鄧小平 1978 年訪日時曾表示：「〔……〕首先承認我們的落後，老老實實承認落後就有希望。再就是善於學習。這次到日本來，就是要向日本請教。」<sup>30</sup>）這個基調一直在一九八〇至九〇年代成為中國主要日本觀。主要的轉變從二〇〇〇年開始，在政治領域研究上，「靖國神社、歷史觀、教科書、軍國主義、小泉」開始慢慢成為主要的研究關鍵字，在二〇〇五年開始有「反日遊行」等等，中日衝突表面化。這與中國二〇〇〇年之後國際地位逐漸攀升，與日本競爭國際地位有關。值得關注的是在中日衝突表面化「之前」，就逐漸有若干研究出現，未來可能藉由觀察「日本研究」動向，來觀察中日關係。另一個並重的是研究「文化社會」等等研究，由於民間社會的來往，日本的動漫流行文化，也逐漸成為關鍵字。

二〇一〇年代以後，特別是釣魚台事件以降，在「國際關係」與「社會經濟」方面，研究增多。社會經濟方面，主要呈現的是「黑暗面」如少子化、地方政治、經濟停滯等等議題；國際關係就更是直接將日本當作對手，甚至台獨、琉球等等敏感問題都直接列入研究範疇。如果跨度大些，讀者很容易看出一些主題、觀點、甚至用詞氣氛大大不同。這顯示中國的日本研究／日本觀點確實發生轉變，從「以日為師」變為「競爭對手」。

下圖為百分比堆疊圖：

30 于俊道：〈生活中的鄧小平〉，《廣安日報》（2019年9月5日），7版。



【圖三】《日本學刊》歷年數量百分比圖（百分比）<sup>31</sup>

## 伍、 結論

本文實踐了將中文文本使用機器學習分類，以達到研究解釋「中國大陸日本觀」之論述成果。就資料處理而言，本文實踐了資料蒐集、資料清洗、文字探勘（尋找關鍵字）、中文斷詞、機器學習分群（LDA 模型）、結合時間與其他標籤製表、以主體知識解釋分群等作為。本文並且指出可使用七分「金融經濟、政治、管理、文學思想、文化社會、國際關係、社會經濟」理解中國大陸近年來日本研究所的研究取向，在歷年的取向中大致維持穩定。但可以細分為「金融經濟／管理／社會經濟」、「政治／國際關係」、「文學思想／文化社會」等組合，由各分群關鍵字（請參看第四節第二小節表），可以看出每個細節關鍵字取向。配合各機構與各年代演變，可以看出「各機構各分類取向」以及「各年代各分類取向」的表格化。

31 資料來源：由作者自行繪製。



本文蒐集了一九九一至二〇一七年，共二二三七篇《日本學刊》的文章，期望回答：「中國社科院日本研究所的日本觀」並回答下列三個問題：

(一) 誰(人/機構)在《日本學刊》上發表？

近十八年中，主要的高產作者有林昶、葉琳、蔣立峰、江瑞平、馮昭奎、王新生、劉江永、金熙德、高洪等。詳細名單請看第四節第一小節表。

(二) 《日本學刊》研究論文有「發表類型」？各類代表性關鍵字有哪些？

依照 LDA 算法，可分為七分群：「金融經濟、政治、管理、文學思想、文化社會、國際關係、社會經濟」。各分群高位關鍵字，請參看第四節第二小節表)。

(三) 各年的「發表取向」有什麼變化？

七分群之佔比為：金融經濟(18.91%)、政治(16.67%)、管理(13.14%)、文學思想(13.99%)、文化社會(12.87%)、國際關係(13.77%)、社會經濟(10.64%)。而「金融經濟/管理/社會經濟」、「政治/國際關係」、「文學思想/文化社會」比率有若干競合關係，可顯示研究取向甚至學科融合的觀點，詳細分析請參閱第四節第四小節。

本文的研究貢獻在於，相較於對單一個人日本觀研究，或是泛論群體的日本觀研究；本文使用自然語言處理與機器學習，能夠擷取一定客觀文本並加以解釋，且其運算結果能夠被客觀實證與重現，為之後的延伸研究做出基礎。而且筆者也實踐了這樣的流程，使得探索性研究能更為便捷，而詮釋性研究也能有所依據。

做為「跨領域新思維」專刊論文，本文試做達成了「資訊領域」與「人文領域」的初級跨域嘗試。這樣的應用與實踐，開啟了更多可能性之視野。吾人可以使用別種語料，如假新聞、法律判決、政治人物言論、社群媒體資料等等，來成就更多詮釋與討論。或是使用不同的演算法，如決策樹、回歸分析等等，來探索更多議題。本文主要的目的，不在聲稱本算法或本語料已經達到完全完整；相反的由於現代科技演變快速，演算法與

資料在寫作期間又有可能更為進步，不管從什麼時間回顧這篇文章，都必然有時間滯後的悖論（Time-lag Dilemma）。也因此，本文期望帶給學界與讀者的是「跨領域新思維」的契機，特別是本刊關心「東亞文明」的可能讀者，在文史哲的領域中，使用數位工具的跨領域研究也可能協助到研究者能夠探索更多未知。比如說本文對大量文字文本的分類，在傳統方法中，必須耗費更大量精力才能做到。

在本研究法之弱點上，如今對「人工智慧與機器學習」的疑惑或評論，同樣也若干適用來批判本文，比如說「演算法本身會有若干歧視，演算出來的結果無法解釋，自動化機器學習法將取代人力甚至取代專家」等等。本文也承認這些批評有其道理，而在具體研究過程中，確實也觀察到了這些現象。不過，根據筆者的經驗，這些恰恰需要研究者具體「跨領域實做」去解決，諸如，研究者自身需要瞭解演算法可能發生的「歧視」，並藉由資料清洗來減少甚至杜絕歧視因子。研究者也需要有主體知識（如本文為日本研究知識）來判斷演算出來的結果是否「離譜」。所以，就機器取代勞動力而言，「有了跨領域能力」的專家，就幾乎很難被取代。因此，綜上所述，反而在本次的研究中，筆者體會到「跨領域新思維」才是解決對於人工智慧的疑慮的可能解方之一。

在細節的研究限制與本文的自身期許上，目前上述流程能夠做到半自動化分群，但分群後之詮釋以及選擇分群數量依舊仰賴與專業研究者之協作，特別在不熟悉的領域中。而且由於統計與抽樣問題，依舊有若干隨機性，在經過語料文本範圍變化（如納入新文獻）或選取不同的門檻值（Threshold），可能會有不同的運算結果，這點也需要專業研究者依照其核心知識來理解詮釋。本文主要核心理念是提出一種演算概念，並以數位工具實踐，協助使用者便於對大量研究文本採取遠程視角，也做為「人工智慧與東亞文明」跨領域研究的一個嘗試成品。

## 引用書目

## 近人文獻

于俊道 YU, Jun-dao

- 2019 〈生活中的鄧小平〉，《廣安日報》（2019年9月5日），7版。  
 “Shenghuozhong de Dengxiaoping [Deng Xiaoping in Daily Life],” *Guang'an Daily*, Sept. 20, 2019, pp. 7.

汪向榮 WANG, Xianzhong

- 1981 〈古代中國人的日本觀〉，《世界歷史》，1981年第5期（1981年5月），頁33-40。  
 “Gudai Zhongguoren de Ribenguan [An Ancient Chinese View of Japan],” *World History*, No. 5 (May, 1981), pp. 33-40.

邵軒磊 SHAO, Hsuan-lei

- 2012 〈中國研究議程之系譜——以日本國際政治學會誌為例〉，《問題與研究》，第51卷第3期（2012年9月），頁23-54。  
 “Zhongguo Yanjiu Yicheng zhi Xipu: yi Riben Guoji Zhengzhi Xuehui zhi Weili [Genealogy of the China Studies Agenda-In the Case of Journal Publications of the Japan Association of International Relations (JAIR)],” *Issues and Studies*, Vol. 51, No. 3 (Sep., 2012), pp. 23-54.

邵軒磊、曾元顯 SHAO, Hsuan-lei & TSENG, Yuan-hsien

- 2018 〈文字探勘技術輔助主題分析——以「中國大陸研究」期刊為例〉，《問題與研究》，第57卷第1期（2018年3月），頁29-62。  
 “Wenzi Tankan Jishu Fuzhu Zhuti Fenxi: yi “Zhongguo Dalu Yanjiu” Qikan Weili [Text Mining-based Topic Analysis: A Case Study of the Journal “Mainland China Studies”],” *Issues and Studies*, Vol. 57 No. 1 (Mar., 2018), pp. 29-62.

侯智元 HOU, Jack

- 2015 〈戴季陶的日本論：思想史視野的觀察〉，《思與言：人文與社會科學雜誌》，第43卷第1期（2015年3月），頁71-98。  
 “Dai Ji-tao de Ribenlun: Sixiangshi Shiye de Guan cha [Dai Ji Tao and His Theory on Japan],” *Thought and Words: Journal of the Humanities and Social Science*, Vol. 43, No. 1 (Mar. 2015), pp. 71-98.

徐興慶 SHYU, Shing-ching

- 2011 〈近代中國知識份子的日本經驗——比較梁啟超、林獻堂、戴季陶的日本觀——〉，《台大日本語文研究》，第21期（2011

年6月），頁179-221。

“Jindai Zhongguo Zhishi Fenzi de Riben Jingyan: Bijiao Liang Qi chao, Lin Xiantang, Dai Jitao de Ribenguan [The modern Chinese scholars’ Japanese experience - comparing Japanese viewpoints among Liang Qichao, Lien Hsien Tang and Dai Qita],” *NTU Studies in Japanese Language and Literature*, No. 21 (June, 2011), pp. 179-221.

張昆、王宇婷 CHANG, Kun & WANG, Yuting

2018 〈中美兩國公眾心中目的日本國家形象——基於2016年中美兩國的全國性民意調查〉，《新聞大學》，2018年第5期（2018年5月），頁124-134。

“Zhong Mei Liangguo Gongzhong Xinzhong Mude Riben Guajia Xing xiang: Jiyu 2016 nian Zhong Mei Liangguo de Quanguoxing Minyi Diaocha [The National Image of Japan in the Public Minds of China and the United States: Based on the national polls between China and the United States in 2016],” *Journalism Bimonthly*, No. 5 (May, 2018), pp. 124-134.

蕭維傑、王維菁 XIAO, Wei-jie & WANG, Wei-ching

2018 〈動態消息演算法、能見度、與新聞經營：以Facebook平台為例〉，《資訊社會研究》，34期（2018年1月），頁63-104。

“Dongtai Xiaoxi Yansuanfa, Nengjiandu, yu Xinwen Jingying: yi Facebook Pingtai Weili [News Algorithms, Visibility, and News Management: Taking the Facebook Platform as an Example],” *Research in the Information Society*, No. 34 (Jan., 2018), pp. 63-104.

BLEI, David M. & NG, Andrew Y. & JORDAN, Michael I.

2003 “Latent Dirichlet Allocation,” *Journal of Machine Learning Research*, Vol. 3, No. 4-5 (Apr., 2003), pp. 993-1022.

BLEI, David M. & JOHN D. Lafferty.

2007 “A Correlated Topic Model of Science,” *Annals of Applied Statistics*, Vol. 1, No. 1 (Jan., 2007), pp. 17-35.

CAMPBELL, David & SHAPIRO, Michael J.

1998 *Moral Spaces* (Minneapolis: University of Minnesota Press, 1998).

CHEN, Titus Chiche & HSU, Chiahao

2018 “Double-Speaking Human Rights: Analyzing Human Rights Conception in Chinese Politics,” *Journal of Contemporary China*, Vol. 27, No. 112 (Feb., 2018), pp. 1-15.

DALLMAYR, Fred

1999 *Bordering Crossing* (Lexington: Lexington Books, 1999).

GRIMMER, Justin & STEWART, Brandon

2013 “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts,” *Political Analysis*, Vol. 21, No. 3 (Jan., 2013), pp. 1-31.

SHAMBAUGH, David

2006 *China Watching* (London: Routledge, 2006).