

深度學「習」：以自然語言處理方法研究 「習近平重要講話」意識形態與時序變化趨勢

邵軒磊

臺北醫學大學醫療暨生物科技法律研究所副教授*

摘要

「習近平之發言作為」成為觀察中共政治情勢的核心指標，其個人思想也成為中共政治的重要動力。本研究試圖回答：「習近平意識形態體系特徵與其思想時序變化」，特別是與毛主義的關係。本研究之方法為「計算政治學 / 計算中國研究」，使用計算機為核心來探索人力難以發現的部分。具體而言，蒐集「習近平系列重要講話資料庫」中之講話文本，作為分析語料。使用程式技術如文字探勘 (text mining)、自然語言處理 (natural language processing) 與深度學習 (deep learning) 演算法，建立分析模型，進行大量自動標註。本研究具體達成自動化機器處理大量語料，並具體實踐其批次流程分析；其次使用自然語言處理抽取其話語特徵；嘗試模型化其意識形態體系。本研究作為一個原型實驗 (prototype experimental) 論文，成功實踐了大量資料蒐集與運算，並且將大規模語料做到「概念分類」，並使用深度學習模型來完成標註與分析。具體而

* 通訊作者，hlshao@tmu.edu.tw。https://orcid.org/0000-0002-7101-5272。本研究為臺北醫學大學新聘教師研究補助 (TMU112-AE1-B30) 暨國家科學及技術委員會「中國研究的知識圖譜：知識萃取、圖譜資料庫、知識生成」(NSTC110-2628-H-038-004-MY4) 研究成果。

言，檢視習近平十年期間（2012年～2021年）的語詞使用以及其意識形態變化。研究發現，習近平意識形態確實有向毛主義螺旋上升的趨勢，對後毛主義提及較為穩定，但鬥爭傾向自「中美貿易戰」後逐漸升高。

關鍵詞：習近平重要講話、自然語言處理、深度學習、計算政治學、文字探勘

* * *

壹、研究背景

2012年11月中共十八屆一中全會後，習近平獲選為中共中央總書記兼任中共中央軍委主席；在2013年3月十二屆的全國人大一次會議，習近平獲選為國家主席兼任國家軍委主席，至今已超過十年。相對於前任「集體領導」的政治模式，一般認為習近平的領導風格更有「個人特色」。如2014年起，推出《習近平談治國理政》，並翻譯為多國文字發行。2016年習近平被定位成中共第五代中央領導集體的核心，並在2017年10月中共十九屆一中全會、2018年3月十三屆全國人大一次會議中連任其職務。

準確來說，「習近平思想」的雛型源自2016年10月27日的中共十八屆六中全會，閉幕後的會議公報提出「習近平總書記系列重要講話精神和治國理政新理念新思想新戰略」的構思。其後，2017年10月18日，由中共十九大報告中提出「習近平新時代中國特色社會主義思想」，將其列為中國共產黨必須長期堅持的指導思想，主要包括「五位一體」、「四個全面」、「中華民族偉大復興」、「人類命運共同體」、「兩個一百年」等，內容涵蓋多種層面，包含政治、經濟、外交、民生、黨政等等。在這樣的累積過程中，習近平的意識形態成爲一種概念集合，因此一般稱爲習近平思想或「習思想」。

隨之，2018年習近平思想正式被寫入中國憲法；同年，全國人大會議通過修憲規定，刪除憲法第七十九條第三款中對國家主席、副主席「連續任職不得超過兩屆」的限制。此舉廣泛被認爲是習近平對於改革開放後中共政治方向的挑戰。更進一步，在政治上中共領導者會將自己的主要政見與貢獻寫入黨章，但近期多不以自身爲名，如江澤民的「三個代表」和胡錦濤的「科學發

展觀」，將「習近平思想」寫入黨章的行為，意味著習近平將與「毛澤東思想」、「鄧小平理論」齊名，並且是繼毛澤東之後，首位仍在世就將姓名列入黨章的領導人，其政治意義不容忽視，可謂「意識形態工程的拐點」（李英明 2018）。

傳統研究中共意識形態，多從符號學（semiotics）的角度，認為中共意識形態往往不真正具備核心價值意義轉變，而更多是藉由某些「能指」（signifier）的使用來達到政治分群效果。以邁斯納（Maurice Meisner）對「馬克思主義中國化」理論為例，毛大量使用烏托邦主義修辭與「國際派」爭奪話語權，在政治操作上區分出「支持毛 / 支持蘇聯」的中共黨員，即是「紅 / 專」衝突，這最終導致毛澤東個人崇拜的政治效果。更進一步，在習近平權威長期化的影響下，使用並引據「習指示」甚至具體字詞之使用，成為中共政治運作的主要動力。

現今研究「習思想」的方法多屬於詮釋學（hermeneutics）或論述研究（discourse studies）；換言之，使用「句式、用語、辭彙」作為能指，來討論其構成及變化。而加上使用計算機甚至人工智慧相關演算法研究文本修辭與規律，已經逐漸成為新興研究方法，如與「計算語言學」類似思路的「計算政治學」（computational politics）。本文即以此研究法為核心，以2012~2021年之「習近平系列重要講話資料庫」中之講話文本，作為基礎的分析語料，企圖運用程式技術的演算方法，以及傳統社會科學的文獻探討與人力識別標記，建立出一套分析習近平思想的深度學習人工智慧建模流程與模型實踐；並且嘗試將中共意識形態與官方語言進行具體化的歸納整理，實踐跨領域中國研究之新方法。企圖瞭解習近平思想的獨特性，以及是否學習前幾位中共領導人的思想道路或理念上有何相似、相異之處。具體而言，本研究使用習近平演講（2012年末~2022年1月5日）作為語料，主要詮釋資料有：日期、全文。本研究的目的並不在討論具體政策（如扶貧、一帶一路）或對政策的臧否，而是討論其意識形態的變化。

使用深度學習研究政治文本的類似著作有：使用政治人物的社群媒體來檢驗貼文的情緒，用以判斷當時的政治氣氛，如支持率（Pota et al. 2018），這個研究較適用於短文本以及零散不同作者文本。另一個相關的研究也是使用推

特 (Kim et al. 2022)。上述針對政治類文本與深度學習的研究，至今多使用社群媒體做為佐證。如前者使用政治人物的推特 (Twitter)，後者使用一般民衆的推特，但上述作者也承認其研究限制在於由於使用預訓練模型，因此對於真正具備解釋力的不確定性以及缺乏對解釋力客觀評估。所以，本文的突破在於使用長文本 (演講文字稿相對於推特貼文)，以及針對單一個體的長期研究。同時本文也使用自造模型方式，即使模型解釋力有限，也能合理的說明其誤差與可能改進的方式。

同時，相較於對於某些僅研究某一時期講話的研究。如使用單次競選活動中的政治人物演講文本比較 (Mahanani et al. 2021) 本研究的特色在於期望研究其「時序」變化，因此使用長時間大量語料的作法。十年間所有的語料共有11,692筆，本研究的主要任務與挑戰也在此。面對龐大的語料，使用全人工閱讀標記再來研究需耗費大量精力。縱使人力閱讀可以使用於某些文本的深入詮釋，筆者自身也曾嘗試深讀少量文本的觀點；但在大量文本時，本研究使用資料分析中的演算法，如文字探勘 (text mining)、自然語言處理 (natural language processing) 與深度學習 (deep learning) 演算法，建立分析模型，再使用模型去大量計算 (標註)，最後再呈現結果，會是因應現代大量資料比較可能實踐的模式。

貳、文獻回顧

關於習近平意識形態走向，目前有若干不同觀點的研究。其中一種是認為習將會是毛式意識形態 (毛主義, Maoism) 的復甦，核心立論是以「習近平中國夢主張」。有學者認為「習的毛主義思想傾向不是簡單地復興毛時代的做法，而是適應實現中國夢的新政治目標的重新挪用...中國夢的思想雖然具有毛時代強烈的民族特色，但也不同於早期的階級鬥爭和革命的總體思想。」 (Yang 2014, 109-113)。在前述文獻中，也強調了習延續前任領導「維穩」思維以及習個人年輕時的文革體驗，是他選擇毛主義的核心原因。另一個相似的觀點指出「習近平的目標是利用毛澤東的遺產來重建政權的合法性，因為它越來越容易受到經濟成長率降低和公衆對腐敗、收入差距和環境污染的憤

怒。」（Zhao 2016, 1168-1193），上述觀點認為這是由於中共之「政權脆弱性」（regime fragility）理論，由這個觀點延伸，就是大家熟悉的中共政治模式，使用意識形態鬥爭與群眾運動，來應對各種政治社會危機，學者張執中（2018）也提出類似說法，本文稍候會檢驗這個觀點。

不過，另一種強調領導繼承的觀點也存在，這意指習近平更多的會遵循鄧小平建立的領導交班制度，並且更多繼受前任領導人意識形態，而非「直接繼承毛思想」。「中國已經出現了三套規範和規則來管理繼任政治，即（1）高層領導人退出權力，（2）繼任者的選擇，以及（3）新領導人的任命。」（Wang and Vangeli 2016, 24-40）。雖然這篇文章最後還是對「繼任制度」表示不確定與擔憂，但基本還是認為「繼任制度與菁英互動」等等給予了若干框架限制，而非習可以任選自己的發展模式。

另一個比較綜合性的角度是認為「中國有某種堅定不移的戰略」，可稱為核心利益，Goldstein在疫情前發表的長文，他指出了中國大戰略（Grand Strategy）並無改變，只是不同時期的領導人表現不同的樣貌，請參見下表一。所以習只是適應了新的環境變化（2020, 164-201）。因此在Goldstein的框架中，習近平應該主要朝向思考並發揚中國核心利益，因此不一定需要遵守毛主席或是後毛主席的方式，而是對於外部國際刺激與內部情勢而發展其意識形態與政策。

表1 1949年至今的中國主要戰略

生存性戰略（1949~1989年） 長遠目標：應對政權面臨的威脅	
三種方法：用以發展、維繫政權生存	
1. 中蘇同盟（毛澤東） 2. 中美合作（毛澤東） 3. 中美結盟（鄧小平）	1. 獲得軍事支持 / 經濟援助 2. 獲得軍事支持 3. 獲得軍事支持 / 經濟合作
中華民族偉大復興（1992年至今） 長遠目標：恢復國際社會強國地位	

三種方法：為復興中華鋪路	
1. 韜光養晦（鄧小平）	1. 為中國崛起累積財富和權力（低調加入現有國際秩序）
2. 和平崛起（江澤民、胡錦濤）	2. 減少國際社會對中國崛起的擔憂（安撫和適應現有國際秩序）
3. 中國夢（習近平）	3. 為中國崛起塑造環境（安撫、改革現有秩序、抵制）

資料來源：Goldstein（2020, 169）。

Goldstein認為即使環境發生變化，中共歷任領導也不會脫離戰略思考，都是服從於中共大戰略的利益而做出權宜表現，自然也沒有路線繼承的問題。不過，這種論述方式把中共領導的接班視為理所當然，而比較少討論「誰能當領導？為什麼？」的議題。比如說，為什麼是習近平接班而非李克強、甚至薄熙來？Goldstein較少討論中共個別菁英在基本的意識形態與利益集團的不同，也不討論意識形態如何影響執政後的領導風格與政治行為模式。

綜上所述，本文主要折衷了Goldstein的觀點，認為研究中共領導者個人意識形態體系並非單純的國家戰略，所以加強對修辭用語的研究。原因在於：

1. 通常中共的政治鬥爭往往向外體現於意識形態路線鬥爭，即使到最後執行的路線相同，在鬥爭當下必須體現其「不同」；具體而言多體現為「用詞」不同。比如說薄熙來事件，當時高層菁英是否要走激進毛主義與溫和改革的路線選擇，與現在的印象大相逕庭，當時習代表的是溫和漸進的中庸路線，在打倒薄路線後，習實質上繼承了毛主義的作法。這也是在某些印象中將2010年代的毛左群眾運動視為習路線的代表的原因，但其實是當時毛左群體是以薄熙來「重慶模式」作為認同對象—這也表示路線鬥爭並不是本質，關鍵還是「權力鬥爭透過意識形態話語」而展現。
2. 但在沒有路線鬥爭的時期，權力還是會透過「語言」來實踐，體現為對象與立場問題，就會往外（外交鬥爭）或往內（階級敵人、疫

情)。這某種程度可以佐證Goldstein的觀察，中共總體而言有一套生存發展戰略，只是對不同的環境下看起來不同。

不過，上述觀點是若干對中共研究解釋，但比較缺少量化指標。這是一個學科方法論本質問題，意識形態研究往往較為宏觀，基礎假設是某個思想家都有一以貫之的思想，所以意識形態研究方有意義。但這樣的說法是否能適用——比如政治家往往必須適應當時政治環境做出決策，甚至是與從前自己的主張相反的決策；或是如前所述，在中共的政治環境，意識形態僅僅是「權力的話語體現」，而不一定是真正的個人信仰。

爲了回答這個理論問題，本研究試著量化習近平在不同時期的「意識形態」，提出「計算政治學」(computational politics)研究法。學者余清祥曾使用類似方式，研究「新聞報導風格」(余清祥、葉昱廷 2020)，並使用可視化技術表現；或是如學者陳至潔使用數量研究中國社交媒體，用以判斷其與政治事件或政治目的之關係(Chen 2022)。因此，本文的理論基礎來自Klimeš and Marinelli (2018)的框架，討論習近平偏向毛主義或後毛主義，Klimeš and Marinelli具體使用了「毛主義」或「後毛主義」的說法。簡言之，強調觀察毛主義(特別是1950~60年代)的革命輸出，內政上表現出集權與個人崇拜，外交上與西方對立。後毛主義則是以「改革開放」政策(1978~)爲核心，其表現與前述相反。「鬥爭」可說是毛主義的延伸，比如說戰狼外交(wolf warrior diplomacy)或是能從「防疫視同作戰、戰勝貧窮」等用語，但直接的演講中直接指明「戰爭或敵意」或是直接從領導人說出較爲少見，因此納入加強語氣用詞，如：「加強、堅持、堅定」爲標註對象，來察覺習近平的鬥爭傾向。Klimeš and Marinelli (2018)論文之結論是習近平不單純屬於兩者(毛澤東時代和後毛澤東時代意識形態)而是對其重申和創新，其證明習近平並未修改共產黨治理基本模式和治理技術的修改。然而，這樣的結論值得更細緻的分析，因此，本文使用深度學習技術作數量分析，主要關心幾個研究指標：

1. 習近平是表現出毛主義，如(Yang 2014, 109-113; Zhao 2016, 1168-1193)的理論？或是表現出比較近似後毛主義(鄧江胡)的意識形態，如(Wang and Vangeli 2016, 24-40)的說法？
2. 依照「危機一鬥爭反應」，當習感覺到「需要鬥爭」時，往往更多會

體現出較強化意識形態的發言。如Goldstein（2000）的說法，是否因為情勢改變，就會影響他表現甚至試圖強化某個方向的意識形態？

當前學界討論習近平思想，往往直覺認為他是走「毛澤東路線」，相似於毛澤東主張要有一套「新秩序」¹；然而也有制度論者認為中共的改革開放時期已經奠定穩定的統治變化，在路徑依賴的前提下，習的意識形態會更多繼承近的領導人。因此，現在就有兩個假說：習近平與毛接近的「復古假說」，以及與鄧江胡接近的「路徑依賴理論」（path-dependent theory）。更進一步，如果習要採取毛主義的論述，在對於既存秩序或是毛意識形態用語上，都會更接近「鬥爭」的方式。細節而言：本文的核心研究議題是檢證一些假設，並配合這些假設延伸出可使用數量化回應的子命題：

命題1：習思想（演講發言）是否較類似毛澤東？（以下簡稱「毛主義」）

命題2：習思想（演講發言）是否較注重經濟或發展？（以下簡稱「後毛主義」）

命題3：習思想（演講發言）中「鬥爭傾向」的語境是否與毛思想的軌跡相似？

命題1較接近於楊國斌（Yang 2014）或趙穗生（Zhao 2016）的理論，認為習因為鬥爭需求或文革經歷，會傾向使用毛意識形態來樹立權威，本文稱為「復古假說」。命題2是王正緒（Wang and Vangeli 2016）的看法，認為較近的領導者設定的框架對現任領導人限制較多，也可能接近「路徑依賴理論」。命題3是，結合「鬥爭理論」與融合Goldstein的說法，認為習要打破若干秩序，就要積極動員黨組織甚至偏向群眾運動的方式，強調危機與動員。

¹ 比如說，在疫情前習近平署名的《共擔時代責任，共促全球發展》就提出：「中國的發展，關鍵在於中國人民在中國共產黨領導下，走出了一條適合中國國情的發展道路。這是一條從本國國情出發確立的道路。這是一條把人民利益放在首位的道路。這是一條改革創新的道路。這是一條在開放中謀求共同發展的道路上。中國的發展是世界的機遇，中國是經濟全球化的受益者，更是貢獻者。中國人民張開雙臂歡迎各國人民搭乘中國發展的快車、便車。」

參、研究設計

茲簡述本文研究設計流程如下：

1. 本文蒐集萬餘篇習近平演講語料（ $n = 11,692$ ），去除「引用報導、僅有圖文、空白」等樣本，並以句子為單位拆分成句語料（ $n = 277,288$ ）。使用句語料是因為往往一「篇」演講中，往往面面顧及，既有社會主義成分也希望經濟發展，甚至也有若干「奮鬥／戰鬥」用語。所以若使用「篇」作為標記單位，這樣標記很有可能並無法真正分辨其意義。
2. 其中，採取間隔抽樣約3,000句給兩位經訓練之研究助理分別標註，扣除太短句或其他無效標註。回收2,509個有效標註。
3. 使用2,509個有效標註，以TextCNN（文字卷積神經網路, Text Convolutional Neural Networks, TextCNN）演算法訓練模型，分別能夠辨識「是否近似毛主義、是否近似後毛主義、是否有鬥爭語氣」。
4. 使用前述模型辨識句語料（ $n = 277,288$ ），並判定前述欄位。
5. 藉由新增的三個欄位判定相互之相關性。
6. 最後，依時間標記合併三個欄位。可以得到「某段時期」習近平講某種語言（毛主義／後毛主義／鬥爭語言）的數量。從而能做出時序分析。

一、資料前處理

在前處理部分本研究蒐集「習近平系列重要講話資料庫」，從2012年末～2022年1月5日的演講，原始樣本有11,747篇。扣除亂碼、字數過少等無效樣本，共得有效樣本11,692篇。隨後，因為有時會有同一則記事重複轉載，常發生於「人民網、新華網、求是」等相關媒體，應為通稿轉載，比如說下列兩篇（粗體為標題）：

向全軍發佈開訓動員令 *新華社北京1月4日電* 中共中央總書記、國家主席、中央軍委主席習近平4日簽署中央軍委2022年1號命令，向全軍發佈開訓動員令。全文如下：我命令：2022年軍事訓練開訓。全軍各級要

堅決貫徹黨中央和中央軍委決策指示，準確把握國家安全和軍事鬥爭形勢變化，緊盯科技之變、戰爭之變、對手之變，大力推進戰訓耦合，大力推進體系練兵，大力推進科技練兵，全面推進軍事訓練轉型升級，練就能戰善戰的精兵勁旅。全軍指戰員要弘揚一不怕苦、二不怕死戰鬥精神，刻苦訓練、科學訓練、安全訓練，練強指揮能力，練好戰鬥本領，練硬戰鬥作風，以昂揚精神面貌和一流練兵成效迎接黨的二十大勝利召開。《人民日報》（2022年01月05日01版）（index：2）

以及

習近平簽署中央軍委2022年1號命令 向全軍發佈開訓動員令 新華社北京1月4日電中共中央總書記、國家主席、中央軍委主席習近平4日簽署中央軍委2022年1號命令，向全軍發佈開訓動員令。全文如下：我命令：2022年軍事訓練開訓。全軍各級要堅決貫徹黨中央和中央軍委決策指示，準確把握國家安全和軍事鬥爭形勢變化，緊盯科技之變、戰爭之變、對手之變，大力推進戰訓耦合，大力推進體系練兵，大力推進科技練兵，全面推進軍事訓練轉型升級，練就能戰善戰的精兵勁旅。全軍指戰員要弘揚一不怕苦、二不怕死戰鬥精神，刻苦訓練、科學訓練、安全訓練，練強指揮能力，練好戰鬥本領，練硬戰鬥作風，以昂揚精神面貌和一流練兵成效迎接黨的二十大勝利召開。（index：5）

這兩者幾乎在文字上相同，因此並不適合重複計算。因此，本研究會使用程式進行比對，留下時間較靠後的部分，剩下9,109篇。隨後，將這個部分分句，以「。；？！」為斷句符號，並將零碎句去除，選取字數大於10字的句子，共277,288句。標註部分，我們使用隨機等距對前述有效文章抽樣，選取時間上平均分配的演講。再將演講提供研究助理進行標註，其標註指南為Lieberthal（1995）所提出概念，再由「毛澤東選集／鄧小平文選」，尋找基本使用詞頻後所得基本關鍵字，再配合前後文脈絡來提出標註指南，舉例如下：

1. 毛主義標註：如文中有明確出現或較為重視社會主義、平等、革命、階級貧窮、黨國意識、第三世界合作、普世價值、國情論等相關內容，便列入標註。比如說：中華優秀傳統文化是中華民族的精神命

脈，是涵養社會主義核心價值觀的重要源泉，也是我們在世界文化激蕩中站穩腳跟的堅實根基。（習近平講故事：以古人之規矩，開自己之生面，人民網—人民日報海外版，2019-08-15, label index：126）或是「闡明這一時期黨面臨的主要任務是，反對帝國主義、封建主義、官僚資本主義，爭取民族獨立、人民解放，為實現中華民族偉大復興創造根本社會條件。（習近平：關於《中共中央關於黨的百年奮鬥重大成就和歷史經驗的決議》的說明，新華網，2021-11-16, label index：133）」

2. 後毛主義標註：是以發展、經濟、韜光養晦、參與國際援助、與大國互助、履行國際責任等為核心的相關句子。比如說「兩國積極展現大國擔當，團結國際社會抗擊疫情，闡述民主和人權的正確內涵，成為踐行真正的多邊主義、維護國際公平正義的中流砥柱。」（習近平：中俄展現大國擔當，成為踐行真正的多邊主義、維護國際公平正義的中流砥柱，新華社微博，2021-12-15, label index：182）或是「中國忠實履行聯合國安理會常任理事國職責和使命，維護聯合國憲章宗旨和原則，維護聯合國在國際事務中的核心作用。」（習近平：中國人民始終踐行多邊主義，中國同聯合國合作日益深化，新華社微博，2021-10-25, label index：251）
3. 鬥爭標註：標註方法為使用鬥爭、軍事用語來描述軍事或非軍事相關對象，特別指出現加強語氣用詞：加強、堅持、堅定、恢復、確保、深化、強勁、動力、新階段、重要時期、提升國力、競爭力、實現偉大夢想等關鍵字眼。實際案例如下，對媒體：在全面建設社會主義現代化國家新征程上，新華社要在黨的領導下，把握正確政治方向，堅定理想信念，堅守人民情懷，賡續紅色血脈，堅持守正創新，加快融合發展，加強對外傳播，努力建成國際一流新型全媒體機構，為實現中華民族偉大復興的中國夢、推動構建人類命運共同體作出新的更大的貢獻。（賡續紅色血脈堅持守正創新努力建成國際一流新型全媒體機構，人民網—人民日報，2021-11-07, label index：751）或是「王勇出席大會並講話，他要求全國應急管理系統堅持以習近平新時代中

國特色社會主義思想為指導，牢記習近平總書記重要訓詞精神，全面落實黨中央、國務院決策部署，堅持人民至上、生命至上，更好統籌發展和安全，盡心竭力做好应急管理各項工作，堅決扛起保民平安、為民造福的神聖職責，努力在新的奮鬥征程上為黨和人民作出更大貢獻、爭取更大光榮。」（習近平會見全國应急管理系統先進模範和消防忠誠衛士表彰大會代表，新華網，2021-11-06, label index : 844）

有時對於人才招募，也會用到「戰場」「戰略」等鬥爭用語，如：做好新時代人才工作，必須堅持黨管人才，堅持面向世界科技前沿、面向經濟主戰場、面向國家重大需求、面向人民生命健康，深入實施新時代人才強國戰略，全方位培養、引進、用好人才，加快建設世界重要人才中心和創新高地，為2035年基本實現社會主義現代化提供人才支撐，為2050年全面建成社會主義現代化強國打好人才基礎。」（深入實施新時代人才強國戰略加快建設世界重要人才中心和創新高地，《求是》202124，2021-12-15, label index : 493）

兩位標註者依照要求各自完成獨立標註後，會進行相互確認。如果標註一致，就維持原標註。若有差異，再由第三研究者進行判斷，最後統整出一個標準標註資料集（golden standard label），並由此做為後續運算的依據。

表2 標準標註資料集數量表

項目	是否毛主義	是否後毛主義	是否鬥爭
負樣本數量	1,663	1,787	2,013
正樣本數量	846	722	496
有效標註總和			2,509

資料來源：作者自行統計整理。

本文標註分類框架理論基礎來自Klimeš and Marinelli（2018）的框架，該文具體使用了「毛主義」與「後毛主義」的說法。簡言之，本文強調毛主義（特別是1950～60年代）的革命輸出，內政上表現出集權與個人崇拜，外交上與西方對立。後毛主義則是以「改革開放」政策（1978～）為核心，其表

現與前述相反。因此，「鬥爭」可說是毛主義的延伸，是與近期中國研究學界中的「戰狼外交」做結合，或是能從「防疫視同作戰、戰勝貧窮」等用語，來察覺習近平的鬥爭傾向。

二、資料處理與運算

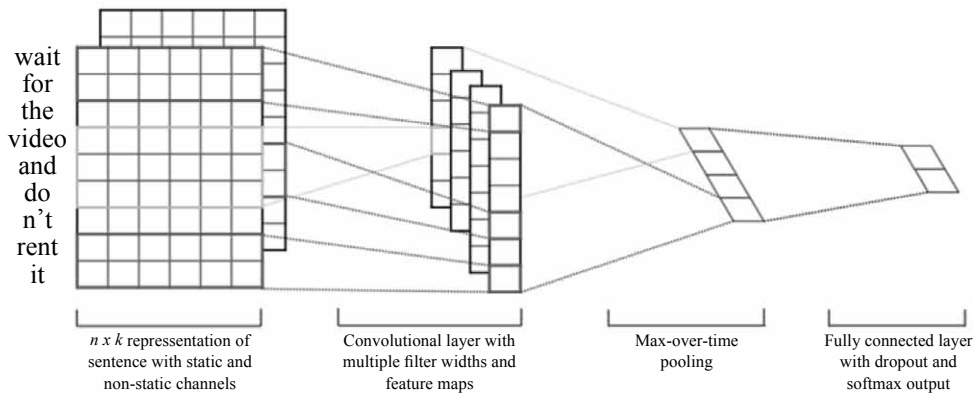
本文將使用深度學習（deep learning）演算法建立模型來判斷意識形態的取向。相對於由研究者的經驗或閱讀出發，本研究傾向使用人工智慧的方式來分析某段話的意識形態。具體而言，使用卷積神經網路神經（Convolutional Neural Networks，以下簡稱：CNN）來「閱讀文字」。這個演算法對人工智慧，特別是電腦視覺方面有不可取代的貢獻，其原理為使用某些核心（kernel）抽取圖片特徵，並藉由某些算式能放大這些特徵，從而辨識物體（Krizhevsky, A., Sutskever, I., and Hinton, G. E. 2012）。電腦視覺是目前人工智慧進展最快也最深入的熱點之一，目前已經進展到「多體連續影像辨識、即時辨識」等領域。（Ren, S., He, K., Girshick, R., and Sun, J. 2015）在CNN的模型與演算過程中，是突破了規則取向（rule-based）的影像辨識，避免了大部分的人工勞動力來描述特徵。所以本文也借鑑了類似的作法，將文本應用CNN的方法，抽取文本特徵並且作文本分類。²

本文使用CNN的特色在於，使用核心（kernel）去卷積（convolute）字詞組合，因此可以一起處理所有的組合，同時也如同前節強調，減少了專家標註的需求。具體而言，若所選取核心大小為5，就會把「為實現中華民族偉大復興的中國夢」一句話，分解成「為實現中華民族偉大」、「實現中華民族偉大復興」、「中華民族偉大復興的」、「民族偉大復興的中國夢」的組合。簡言之，使用n-gram的概念，將五個字詞一起讀，可將前後文的連結意義讀取進去，實現了對於順序型資料的意義與特徵抽取。CNN演算法

² 本模型使用之相關參數如下：segmentation=jieba, word_vector=BertEmbedding, word vector, dimension=100, window=5, text category= binary, batch size=128, iteration round=1000, discard rate=0.5, learning rate=0.001, vocabulary=wordembedding, activation function=relu, optimizer=Adam。

等於幾乎是將「所有字詞組合特徵」都抽取出來，並結合其相互關係來強化特徵；因此，在良好的運算結構設計下，就能夠順利辨別其特色。之後，再使用深度類神經網路強化特徵權重（如下圖所示）。換言之，這些抽取特徵的過程都能夠大量自動化，而不需要專家一一描述關鍵字；但代價是電腦算力需求與模型建設工程。

圖1 卷積神經網路結構圖



資料來源：Kim, Y. (2014)。

在本研究中，本文將輸入「一段習近平演講語句」，經由模型給出標註。再將「預測值」與「實際值（人類標註值）」做比較，即可知道模型之正確率。

肆、數據結果呈現

一、自動標記模型訓練成果

本次採分別訓練的三個模型，其預測正確率約為0.78~0.85。其詳細效能如下表所示：

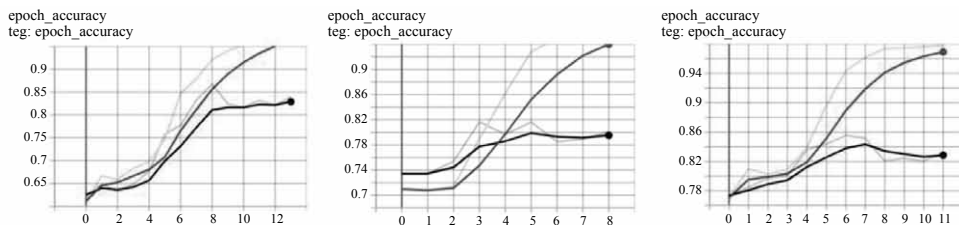
表3 本研究自動標註模型效能表

	precision	recall	f1-score
ifMaoism			
0	0.8545	0.916	0.8842
1	0.8073	0.6929	0.7458
accuracy			0.8408
ifPostMaoism			
0	0.8345	0.8625	0.8483
1	0.6263	0.5741	0.599
accuracy			0.7798
ifFight			
0	0.8769	0.9669	0.9197
1	0.7727	0.4533	0.5714
accuracy			0.8647

註：ifPostMao=後毛主義模型，ifFight=鬥爭傾向模型，ifMaoism=毛主義模型。postMao_sum=後毛主義。f1-score結合了模型的精確度（Precision）和召回率（Recall）的評估指標，用於衡量分類模型的性能。精確度表示模型預測為正類的樣本中實際為正類的比例，召回率表示實際為正類的樣本中被模型正確預測為正類的比例。F1值則是精確度和召回率的調和平均，為了防止不均衡數據造成誤差，將兩者結合起來，能夠更全面地評估模型的性能。accuracy=模型預設正確率。

下圖呈現訓練紀錄：

圖2 深度學習模型訓練紀錄組圖



註：左：model_ifMaoism；中：model_ifpostMaoism；右：model_ifFight。橫軸為訓練次數，縱軸為正確率。淺色線為訓練組正確率，深色線為驗證組正確率。當正確率逐步提升，我們能夠確定模型訓練成功。即使訓練組可能持續上升，但因為語料特性驗證組正確率會到達極限。當開始持平時，實驗者會停止訓練。

隨後，我們再使用這三個模型，對所有「句子」（ $n=277,288$ ）做自動標註。

二、標計後數值分析

（一）標註數值敘述統計

若以句為單位進行標註，我們可以得到每一句使用的意識形態標註數值。

表4 習語料每句意識形態傾向

	Maoism	postMaoism	ifFight
無標註	171,734	218,945	242,263
正向標註	105,554	58,343	35,025
正向標註 (%)	38.07	21.04	12.63
總和			277,288

註：PostMaoism=後毛主義相關，ifFight=鬥爭相關，Maoism=毛主義相關。

從表四可以看出句子中，有約38%的比率會出現類似毛主義的論述。而傾向比較後毛主義領導人的論述，約有21%。

第二階段的處理，是將句子重組成為原來演講，可以得到演講的傾向。由於合併時會隨著文章的「句子數」而可能發生差，比如說某個演說稿特別長，這樣的話，「比率（文章內某種標註句數量 / 文章內總句數量）」更能反映出其意義。因此，我們重新合併時使用「比率」與「總和」兩種觀察指標。不過兩種指標都顯示習的用語中，更接近毛主義。

表5 習語料每篇平均意識形態傾向

	平均數	標準差	最小值	中位數	最大值
postMao_ave	0.2	0.172	0	0.174	1
ifFight_ave	0.105	0.126	0	0.067	1
Maoism_ave	0.341	0.248	0	0.301	1

註：每篇文章之postMao_ave=後毛主義平均值，ifFight_ave=鬥爭傾向平均值，Maoism_ave=毛主義平均值。標註前9,109篇是去除重複值（如新華網或人民網重複刊登）後的有效樣本。在拆分後自動標註後，會去掉掉零碎句子（字數小於10的句子），進行自動標註再重新重組成「文章」。如果這篇文章本身句子過短，就會在這個階段被剔除。因此最後有效呈現運算的是7,570篇。

上表顯示，如果用比率表現，每篇演講中約有34.1%的部分會出現毛主義語句，而20.0%會出現後毛主義句子，10.5%會出現鬥爭句子（其餘是不歸屬於任何一個分類的用語）。這些數量會隨著具體演講對象或文本長短而有差異，不過相對來看是支持習近平的「毛主義」傾向。

（二）命題驗證

- 命題1：習思想（演講發言）是否較類似毛澤東？抑或
- 命題2：習思想（演講發言）是否較注重經濟或發展？

表4、表5數據指出，習演講發言中毛主義絕對數量與佔比都是較多的，因此本文支持「復古假說」（命題1）。

- 命題3：習思想（演講發言）中「鬥爭傾向」的語境是否與毛思想的軌跡相似？

習思想中提到鬥爭的時候通常為政治目標與黨建和危機處理，典型如：扶貧攻堅戰、反腐鬥爭、疫情³及動員處理地方危機時（水災⁴、地震⁵）、黨史

³ 習近平：在統籌推進新冠肺炎疫情防控和經濟社會發展工作部署會議上的講話 同志們：今天，我們召開統籌推進新冠肺炎疫情防控和經濟社會發展工作部署會議，主要是分析新冠肺炎疫情防控和經濟社會發展工作。下面，我講幾點意見。一、關於前一段疫情防控工作新冠肺炎疫情發生後，黨中央高度重視，迅速作出部署，全面加強對疫情防控的集中統一領導。1月7日，我主持召開中央政治局常委會議時，就對做好疫情防控工作提出了要求。1月20日，我專門就疫情防控工作作出指示，要求各級黨委和政府及有關部門把人民群眾生命安全和身體健康放在第一位，採取切實有效措施，堅決遏制疫情蔓延勢頭。大年初一，我主持召開中央政治局常委會議，對疫情防控工作進行再研究、再部署、再動員…（index: 2145）。

⁴ 始終把保障人民群眾生命財產安全放在第一位 抓細抓實各項防汛救災措施 中共中央總書記、國家主席、中央軍委主席習近平對防汛救災工作作出重要指示。習近平指出，近日，河南等地持續遭遇強降雨，鄭州等城市發生嚴重內澇，一些河流出現超警水位，個別水庫潰壩，部分鐵路停運、航班取消，造成重大人員傷亡和財產損失，防汛形勢十分嚴峻。習近平強調，當前已進入防汛關鍵期，各級領導幹部要始終把保障人民群眾生命財產安全放在第一位，身先士卒、靠前指揮，迅速組織力量防汛救災，妥善安置受災群眾，嚴防次生災害，最大限度減少人員傷亡和財產損失。解放軍和武警部隊要積極協助地方開展搶險救災工作…（2021年07月22日01版）（index: 504）。

⁵ 習近平總書記打電話對抗震救災工作作出重要指示 22日晚10時5分，中共中央總書記、國家主席、中央軍委主席習近平，與四川省委書記、省抗震救災指揮部指揮長王東明通電話，對四川蘆山地震災區人民表示親切問候，對奮戰在抗震救災一線的解放軍指戰員、武警部隊官兵、公安幹警和各類救援隊伍表示親切問候和感謝。在聽取王東明簡要彙報後，習近平總書記對四川抗震救災工作給予充分肯定，對進一步做好抗

回顧等。毛意識形態往往強調「革命與鬥爭」（D’Mello 2009），因此我們將逐篇含有毛主義／後毛主義／鬥爭思想三者做出相關分析。

表6 習語料中鬥爭與毛思想／後毛思想相關表

correlation	postMao_ave	ifFight_ave	ifMao_ave
postMao_ave			
ifFight_ave	-0.0115		
ifMao_ave	-0.0548***	0.5527***	

註：***: p-value < 0.001。每篇文章之postMao_ave=後毛主義平均值，ifFight=鬥爭傾向平均值，Maoism_ave=毛主義平均值。

毛思想與後毛思想有負相關，這是比較合理的；同時也可以看出毛思想與「鬥爭用語」的文章相關性有0.5527，中度正相關。而後毛思想與「鬥爭用語」是-0.0548，微小負相關。這與學者張執中（2018）預期「習近平使用危機模式逐漸鞏固權力」相符。比如說在「香港抗議、中美貿易戰、新冠疫情」等危機中，藉由動員等應對策略逐步強化個人的政治權力和領導地位。或是藉由改革和整頓重新整理人事，增加「工作小組」重塑權力結構，強調對黨內外的統一領導。以及加強意識形態控制，強化對媒體、教育和文化領域的控制，以塑造對個人或黨派的支持和忠誠。

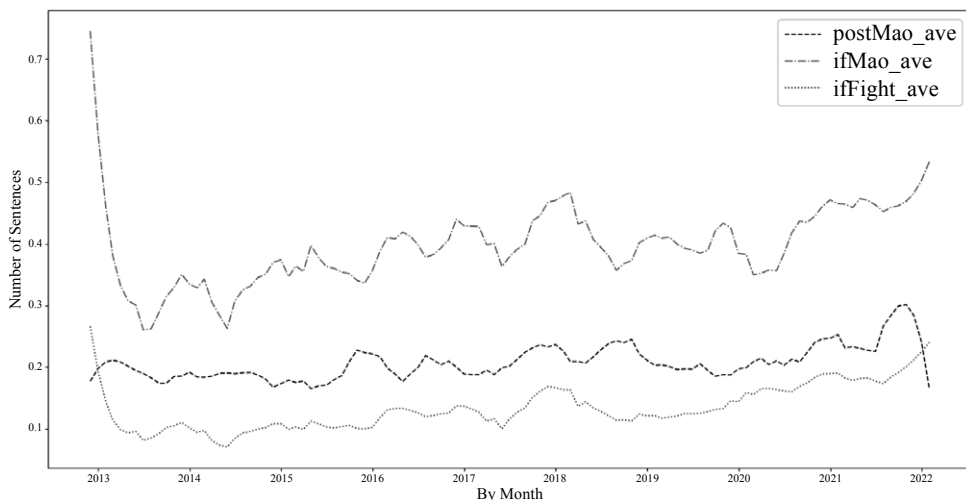
三、時序變化

（一）時序分析

在時序分析上，爲了避免某些時候「發表比較密集」，也由於長達10年的期間，本研究使用「月」來整合時序資料。由於某些月份（通常3月、5月、10月、11月）因爲事件而特別多講話機會，或是發表了一篇特別長的講話。這樣會容易被這樣的演講帶偏離數值，因此會使用一個比例時序圖，即將「所有句子數」作爲分母，某一標記值總數量做爲分子，看每個月的時序分佈。

震救災工作提出了明確要求。他說，黨中央、國務院對蘆山4·20地震抗震救災工作高度重視，第一時間緊急動員和部署…（index: 11289）。

圖3 月平均比例時序圖（2012~2022）



註：每篇文章之postMao_ave=後毛主義平均值，ifFight_ave=鬥爭傾向平均值，Maoism_ave=毛主義平均值。

表7 習語料逐月鬥爭、毛思想 / 後毛思想平均數量高點列表

Month_Ave	Peak_Period
Maoism	2013/12, 2015/6, 2016/8, 2016/12, 2018/3, 2020/1, 2022/1
PostMaoism	2015/11, 2016/6, 2017/11, 2018/9, 2021/11
Fight	2017/11, 2018中期後持續上升

註：PostMaoism=後毛主義相關，Fight_ave=鬥爭相關，Maoism=毛主義相關。

從比例可以看出在2017年10月（第二任期）為重要分界，前期大約呈現半年一個循環的趨勢，其後就比較紊亂。尤其在中美貿易戰之後，如「戰勝疫情」使得鬥爭用語持續上升。

伍、研究成果

一、研究貢獻與假說討論

本研究作為一個原型實驗（prototype experimental）論文，成功實踐了大量資料蒐集與運算，並且將大規模語料做到「概念分類」，並使用深度學習模型來完成標註與新研究。並且完成了用來檢證了幾個對習近平意識形態的兩個命題——「復古假說」/「制度依賴理論」。細論如下：

命題1：習思想（演講發言）是否較類似毛澤東？

從圖3來看，習思想中提起毛主義的次數相較後毛主義高上不少，因此我們可以從數量上來實證，確實習思想比較接近毛思想。其中一個可能的解釋是後毛時期的鄧江胡三任造成了龐大黨官僚組織，這對習初期作為弱勢總書記是不利的，而毛的政治生涯也曾面臨過這樣的問題，因此參考與效法毛主義似乎成為他合適的選擇。習近平的意識形態在這十餘年間也可能轉型，因應不同場合需求，習近平同時對毛澤東時代和後毛澤東時代意識形進行重申和創新，所以毛主義與後毛主義並非相互拮抗的關係，但在第二任期就開始有較為明顯相反的趨勢。這可能是觀察時間點與選擇研究材料的影響。所以本文的目的在於建立一個研究模式。從而能夠提供後續的研究者判斷的基礎，而非各說各話。從圖四來看，每年三月兩會時期，社會主義的傾向就會出現高峰。而成螺旋上升的情況。

更細部而言，由可以看出，毛主義會在特定的時間內出現持續攀升、下滑的趨勢，在與其相關知識背景的結合下，可以判斷出習近平思想在針對國內政治、黨建議題的講話上，會有毛主義上升的情況。例如，在2014年中～2015年中，毛主義明顯持續攀升，並在2014年末～2015年初達到高峰，便與習近平在2013年3月14日正式擔任國家主席有關。開啓新時代的序幕之際，為求穩定黨內的統治地位，除了推動黨內的改革之外，也會在談話中不斷提及堅持馬克思主義、黨國意識、人民、民族精神等字句，如在2013年3月17日第十二屆全國人民代表大會的第一次會議講話上，習近平強調「中國共產黨的領導和團

結、建設中國特色社會主義偉大事業」；「實現中國夢必須走中國道路、弘揚中國精神、凝聚中國力量、中國夢是人民的夢」；「堅持黨的領導、鞏固和發展愛國統一戰線」等，以意識形態的策略來加強自身統治的正當性，並陸續推出一系列習近平思想。由此可以印證，習近平在維護黨內地位以及推行內政改革時，往往會與毛主義接近。而國際關係與事務的變化也會對毛主義的傾向產生影響，例如，2015年至2015年中後的曲線下滑，可能與該年9月習近平訪問美國並參與聯合國成立70周年峰會有關。此次出訪習近平提倡不衝突、不對抗、相互尊重、合作共贏的「新型大國關係」，藉此增進美中兩國的信任與合作。此外，2015年同時也是中國積極推動一帶一路與亞投行（AIIB）的重要日程，因此，在這個階段基於降低國際對於中國崛起的猜忌，以及化解美國對於中國挑戰霸權秩序的疑慮考量下，毛主義的比例在國際舞臺上往往會有顯著的下降。

這個趨勢在川普時期與習第二任期時更為明顯看出2015年中後～2016年初的毛主義攀升。從2015年10月底的五中全會；2016年1月亞投行的正式開幕以及2016為十三五計畫執行的起始年。其中十三五計劃的重要實施包括：一帶一路、2020全面脫貧、全面建成小康社會、中國製造2025等。亞投行則是自2015年6月，便有50個亞投行創始會員國共同簽署《亞洲基礎設施投資銀行協定》，並陸續增加簽署國，直至2015年12月該協定達成生效。2016年1月習近平正式發表亞投行成立的重要談話，並提倡「開放包容的區域主義、多邊主義」。此外，2016年3月亦有「人民代表大會」和「中國人民政治協商會議」的兩會召開；7月中國共產黨成立95周年，因此毛主義下滑速度趨緩，直至2016年末，因美國和台灣分別由川普、蔡英文贏得選舉後才急遽轉變。2020年～2021年的持續高峰，不僅是因為新冠疫情爆發，社會需要精神的支援，也有2020達成全面小康（十三五計畫目標），以及2021為中共建黨百年的因素存在。

自2017年十九大後，習思想進入黨章，並且於隔年即取消國家主席連任限制，自這個時間點開始習提及毛主義的次數居高不下，整體而言是螺旋上升的狀態。其中有兩個時期是值得注意的，一是2019年後的「中美貿易戰」時毛主義比例減少，這反映了因為受到美國的挑戰，而中共感覺無法控制外部，因

此顯示權力不穩的狀態；二是2020年「疫情」，使得中共使用動員方式加強對內部的控制，因此有毛主義增加的傾向。

命題2：習思想（演講發言）是否較注重經濟或發展？

習提及鄧江胡路線的次數明顯不如他的前任（圖3），從習的行為便可解答，他有意在政治上傾向毛主義，而較不遵循鄧小平留給黨的制度框架如：國家主席連任限制、有意的黨政分離、隔代指定接班人、集體領導等規則，但在經濟政策上卻又不大敢完全背離鄧路線，而是在毛鄧之間遊離，相對地，在鄧江胡在任時不具備這樣的傾向，因此習實際上並不如接近毛那樣接近前任幾位領導人。

一般而言，我們可以從圖四發現毛主義與後毛主義多數呈現相反趨勢關係。但習近平在國際會議或與他國政要會談的場合，通常談話性質較為和緩，而是提倡和平、共贏、合作、互利的價值。例如2015年習近平訪美後，於9月26日前往紐約聯合國總部，宣佈成立南南合作支援基金；9月28日出席聯合國成立70周年，並在峰會發表談話，內容提到中國在同月3日舉行的紀念中國人民抗日戰爭暨世界反法西斯戰爭勝利70周年大會，以及主張構建以合作為核心的新型國際關係，建設「人類共同體」，反對國家之間欺負弱小、不平等的行為，反對任何國家干涉、侵犯他國內政與領土主權。所以在習出席國際場合，希望改善對外關係或是國內外經濟合作、議題時，後毛主義有時也會平衡毛主義的作用，當國內外輿論質疑是否要拋棄鄧路線時便會釋出來消除國內外懷疑；例如：金融、房地產改革、2020雙循環經濟政策等議題。

命題3：習思想（演講發言）中「鬥爭傾向」的語境是否與毛思想的軌跡相似？

在資料後期，習用語中出現鬥爭的情況大量上升，其因為習自身的意識形態傾向的用語有關，傳統上；出現鬥爭傾向的用語往往出現於政治路線、權力的鬥爭與危機處理時的動員，因此在圖四中習上臺的第一任期雖然有反覆，但鬥爭緩緩上升，在2017年中期起急速上升。如上所說，都是在進行路線與權力鬥爭，之後隨著情況穩定後呈現較為平緩但時有上升又下降的狀況，隨後又

逢十九大，便又上升。在習時代被傳統定義所匡列範疇的便是如：疫情時的全國動員、薄熙來的權力鬥爭、徐才厚與郭伯雄的倒臺，而非傳統定義範疇的便是針對歐美外交行動的不友好而採取的反擊措辭或是宣傳中國民族主義時的用語，典型的如：中華民族的偉大復興、中國夢、中美阿拉斯加峰會、中澳外交惡化等。總而言之，鬥爭的軌跡大致上在習時代有此規律：習新政策時有一波上升期，而下降則取決於習的權力穩固時間，之後便隨著國內外政治情況再決定上升與否，發生危機或是危及政權時上升，反之則平穩或是下降，若發生習權力受威脅則上升，反之則下降，周而復始為一循環。

特別是「鬥爭項目在2014年前期總量上升但比例下降」，估計是因為開始有某些事情的徵兆出現（如昆明301事件），但同時有更多的其他事務需要處理。2014年是進入習李體制進入適應期，也是「一帶一路」等向外積極外交政策的時期。因此，即使在某些事件上升，但主旋律會被「一帶一路」的主旋律帶著走，因此會出現這樣的差異。

從圖四來看，鬥爭可說是大致與毛主義掛勾。然而，因後毛主義的平衡作用，導致曲線起伏較於平緩（跟毛主義相比）。誠如上述所言，正因為毛主義大多是針對國內黨政議題採取的強硬手段或是執行口號，因此鬥爭意味十足。像是：奮鬥目標、確保目標達成、恢復國際地位、實現中國夢、偉大復興等。其中，在提到中國夢時也會和過去的「百年國恥」掛勾，而產生一種現代中國超越過去落後中國的感覺，並以此建立為人民美好未來想像的藍圖。

從圖四來看，尤其在「中美貿易戰」後，鬥爭性質是有上升的趨勢。例如川普上任後的舉措導致中美關係惡化，中國為緩解國內對貿易的動盪與不安，整體的措辭趨於強硬。表現在疫情期間，甚至「後疫情」時期，貌似鬥爭語言已經「不可逆」的狀態。綜上所述，本文支持楊國斌（Yang 2014）或趙穗生（Zhao 2016）的理論，習越來越趨近毛的行為模式。

在對每個演講稿觀察時，筆者稍微發現對外賓（如外國元首使節、國際組織）有較為緩和的趨勢，而對「疫情、黨員」時，會較為嚴厲的批判，展現出鬥爭的傾向。可能的解釋是中共黨文化的「內外有別」傾向（邵軒磊 2019），不過總體而言，還需要後續發展更多有系統的研究。

二、研究限制

本研究之特色與限制都來自於人工智慧標註模型，由於本次的資料集，人工智慧模型在訓練的過程中難達到完美，正確率約在79%~85%之間。據此標註的全樣本資料，也一定有若干誤差範圍。因此在後續的推論過程中，我們要謹記這個誤差範圍始終存在。不過由於樣本數量較大，所以在大數法則的作用下，若干程度可以消滅誤差影響。同時，也必須思考未來如何提升本研究模型的正確性。

使用句標註的嘗試也感謝學術前輩的指引，唯在運算過程中「句」的資訊流失也是無法避免，比如說過短句子是否應該列入計算。或是有些詞語用以作為標註對象語句的合理性，是因為考慮到現實數據樣本（尤其是文字樣本）確實無法完全找到明確規範。本文使用一個暫訂的流程，有待被更多的研究超越或揚棄。而「可視化」部分本身就是一種將不確定概念轉化的過程，這也有若干可能改進的空間；但能夠具像化「意識形態」，而且一些意識形態循環的規律，並具體觀察到「中國內政外交關係的鬥爭詞語」的上升趨勢，是本文目前可以做到的貢獻。

無寧說，筆者寧願強調原型實驗（prototype experimental）的原創性：其一是本研究之模型訓練停留在實驗層級；如果邁向具體使用，應該設法提高模型正確率。可能有更優越的演算法，不過目前尚未被研究者發現。二是在標註品質上，本研究提供研究助理等級的標註與理解，如果真正具體使用，應投入更多專家人力來進行標註。這兩者都可能在資源投入或者擴大研究團隊規模再來更深入回答。

三、研究貢獻與未來展望

在學術上，本研究使用數據方式回答了意識形態問題，使得意識形態問題（習是否為毛主義者）可以有一個基礎的數據解答。其次，本研究室亦提供了相關類似題目的研究流程設計，這樣的設計使得可以控制研究成本，亦可省下重複作業的時間，此外更可以以此模型做出中共中央及習近平個人思考方向與政策趨勢的預測，更可以用這樣的數據方式來研究其他中共領導人。

有學者建議未來可以具體觀察「演講對象」是否有區別，這是非常好的發展方向，依照傳播／宣傳理論上應該會有差異，筆者整理研究資料與閱讀每篇演說資料時也發現若干「不同」；以及建議使用「段落」為分析單位也可能有發展空間，確實值得嘗試。另外，在文本分類演算法上，有使用BERT與CNN的Chen et al. (2022) 以及Zhao et al. (2022) 使用WTL-CNN的成功案例，皆是值得嘗試與檢驗的。不過限於篇幅與本文主題，期望將來能有更多技術合作或是未來延伸研究的機會。

(收件：112年3月25日，接受：113年6月19日)

Deep Learning “Xi”: Using Natural Language Processing Important Ideology Model and Timeseries Trends in “Xi Jinping’s Speech”

Hsuan-lei Shao **

Associate Professor
Graduate Institute of Medical and Biotechnological Law,
Taipei Medical University

Abstract

Xi Jinping’s activities have become the core focus of the CCP’s political landscape, and his ideology has emerged as a significant driving force in Chinese politics. This study aims to answer the question: “What are the characteristics of Xi Jinping’s ideological system and its chronological evolution?” particularly in relation to Maoism. The methodology of this research is rooted in “computational politics/computational Chinese studies,” utilizing computational methods to explore aspects that are difficult to uncover manually. Specifically, the study collects speech texts from the “Xi Jinping Series of Important Speeches Database” as analytical material. By employing programming techniques such as text mining, natural language processing (NLP), and deep learning algorithms, the study builds analytical models for extensive automated annotation.

This research achieves two primary goals: first, it automates the processing of large volumes of textual data and implements batch process analysis;

** Corresponding Author: hlshao@tmu.edu.tw, This research was supported by the New Faculty Research Grant of Taipei Medical University (TMU112-AE1-B30) and the National Science and Technology Council’s project “Knowledge Map of China Studies: Knowledge Extraction, Map Database, and Knowledge Generation” (NSTC110-2628-H-038-004-MY4).

second, it uses NLP to extract discourse features and attempts to model Xi Jinping's ideological system. As a prototype experimental paper, it successfully implements large-scale data collection and computation, achieving “concept classification” of extensive corpora and utilizing deep learning models for annotation and new research. Specifically, the study examines the vocabulary usage and ideological changes over Xi Jinping's ten-year period (2012-2021). The findings indicate that Xi Jinping's ideology indeed shows a spiral upward trend towards Maoism, with a relatively stable reference to post-Maoism, but with an increasing tendency towards struggle following the “China-U.S. trade war.”

Keywords: Xi Jinping's Important Speech, Natural Language Processing, Deep Learning, Computational Politics, Text-Mining

參考文獻

- 邵軒磊，2019，〈機器學「習」：以文字探勘法探索習近平時期之大外宣戰略〉，《中國大陸研究》，62（4）：133-157。Shao, Hsuan-lei. 2019. "Ji qi xue xi: yi wen zi tan kan fa tan suo xi jin ping shi qi zhi da wai xuan zhan lue" [Machine Learning: An Application of Text Mining to Xi's Grand External Propaganda Strategy]. *Mainland China Studies*, 62 (4): 133-157.
- 余清祥、葉昱廷，2020，〈以文字探勘技術分析臺灣四大報文字風格〉，《數位典藏與數位人文》，6：67-94。Yue Ching-syang Jack, Yu-Ting Ye. 2020. "Yi wen zi tan kan ji shu fen xi tai wan si da bao wen zi feng ge" [A Study of Text Mining on Taiwan's Newspapers]. *Journal of Digital Archives and Digital*, 6: 67-94.
- 李英明，2017，〈習近平思想：新的中共意識形態模式的興起〉，《展望與探索月刊》，15（9）：8-13。Lee, In-ming. 2017. "Xi Jin-ping si xiang: xin de zhong gong yi shi xing tai mo shi de xing qi" [Xi Jin-ping Thought: The Rise of a New CCP Ideological Model]. *Prospect and Exploration*, 15 (9): 8-13.
- 張執中，2018，〈黨國意識形態的權力與制度：習近平時期的觀察〉，《展望與探索月刊》，16（4）：71-88。Chang, Chih-chong. 2018. "Dang guo yi shi xing tai de quan li yu zhi du: Xi Jin-ping shi qi de guan cha" [An Observation of Power and System of Party-State Ideology in Xi Era]. *Prospect and Exploration*, 16 (4): 71-88.
- Chen, Titus C. 2022. *The Making of a Neo-Propaganda State: China's Social Media under Xi Jinping*. Brill.
- Chen, Xin-ying, Peimin Cong, and Shuo Lu. 2022. "A Long-text Classification Method of Chinese News Based on BERT and CNN." *IEEE Access*, 10: 34046-34057.
- D'Mello Bernard. 2009. "What is Maoism?" *Economic and Political Weekly*, 2009 (1): 39-48.
- Goldstein, Avery. 2020. "China's Grand Strategy under Xi Jinping: Reassurance,

- Reform, and Resistance.” *International Security*, 45 (1): 164-201.
- Kim, Daesik, Chung Joo Chung, and Kihong Eom. 2022. “Measuring Online Public Opinion for Decision Making: Application of Deep Learning on Political Context.” *Sustainability*, 14 (7): 4113.
- Kim, Yoon. 2014. “Convolutional Neural Networks for Sentence Classification.” *arXiv preprint*, arXiv: 1408.5882.
- Klimeš, Ondřej, and Maurizio Marinelli. 2018. “Introduction: Ideology, Propaganda, and Political Discourse in the Xi Jinping Era.” *Journal of Chinese Political Science*, 23: 313-322.
- Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2012. “ImageNet Classification with Deep Convolutional Neural Networks.” *Advances in Neural Information Processing Systems*, 25.
- Lieberthal, Kenneth. 1995. *Governing China: From Revolution Through Reform*. New York: W. W. Norton & Company.
- Mahanani, Rista Yulisetya, Harun Joko Prayitno, and Abdul Ngalim. 2022. “Assertiveness around the 2019’s Presidential Candidacy Discourse on Social Media: A Sociopragmatic Study.” *LiNGUA*, 16 (1): 139-150.
- Pota, Marco, et al. 2018. “A Subword-Based Deep Learning Approach for Sentiment Analysis of Political Tweets.” 2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA). IEEE.
- Ren, Shaoqing, et al. 2015. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.” *Advances in Neural Information Processing Systems*, 28.
- Wang, Zhang-xu, and Anastas Vangeli. 2016. “The Rules and Norms of Leadership Succession in China: From Deng Xiaoping to Xi Jinping and Beyond.” *The China Journal*, 76 (1): 24-40.
- Yang, Guo-bin. 2014. “The Return of Ideology and the Future of Chinese Internet Policy.” *Critical Studies in Media Communication*, 31 (2): 109-113.

Zhao, Sui-sheng. 2016. "The Ideological Campaign in Xi's China: Rebuilding Regime Legitimacy." *Asian Survey*, 56 (6): 1168-1193.

Zhao, Wei-dong, et al. 2022. "WTL-CNN: A News Text Classification Method of Convolutional Neural Network Based on Weighted Word Embedding." *Connection Science*, 34 (1): 2291-2312.