



數位人文視角下清代黃冊檔案的 犯罪類型、動機與時空分布

A Digital Humanities Approach to Crime Records in Qing Dynasty Yellow Registers: Typologies, Motives, and Spatiotemporal Patterns

邵 僅* 林巧敏** 

Chin Shao * Chiao-Min Lin **

【摘要】

本研究以清代黃冊檔案中之刑事案件紀錄為分析對象，結合生成式人工智慧（Generative AI）、Python 與 DocuSky 等數位人文工具，探討犯罪類型、犯罪動機與案件的時空分布特徵。研究採用中央研究院近代史研究所典藏之 1,745 筆漢文黃冊檔案，透過文本預處理與語義分析、自動化標註、犯罪類型與動機分類，並進行詞頻統計及地理資訊視覺化處理。結果顯示，殺人與傷害等暴力犯罪為最主要的案件類型，「家庭糾紛」、「感情糾紛」與「偶發衝突」為常見犯罪動機，並具有高度的時代與空間集中性；犯罪紀錄高峰期集

* 國立政治大學圖書資訊與檔案學研究所博士生
Ph. D. Student, Graduate Institute of Library, Information and Archival Studies,
National Chengchi University
E-mail: 114155501@g.nccu.edu.tw

** 國立政治大學圖書資訊與檔案學研究所特聘教授
Distinguished Professor, Graduate Institute of Library, Information and Archival
Studies, National Chengchi University
ORCID 0000-0002-9309-9884
通訊作者 Principal author for all correspondence E-mail: cmlin@nccu.edu.tw

中於康熙中後期與乾隆末期，地理分布則以山東、江南與廣東等人口稠密地區為高頻區。本研究旨在以數位人文方法分析非結構化的歷史檔案資料，探究將數位方法應用於犯罪動機語義判讀與司法歷史研究的潛力，期許可提供應用史料的不同研究取徑。

關鍵詞

數位人文 清代黃冊 戶籍檔案 犯罪紀錄 文本探勘

【 Abstract 】

This study analyzes criminal case records found in Qing Dynasty Yellow Registers, employing a digital humanities approach that integrates generative artificial intelligence (GAI), Python programming, and the DocuSky platform. Drawing upon 1,745 Chinese-language Yellow Registers entries preserved by the Institute of Modern History, Academia Sinica, the research applies text preprocessing and semantic analysis to enable automated annotation and classification of crime types and motives, complemented by word frequency statistics and geospatial visualization. Findings indicate that violent crimes such as homicide and assault constitute the predominant categories, with “domestic disputes,” “romantic conflicts,” and “spontaneous altercations” emerging as the most frequent motives—each displaying notable temporal and spatial concentration. Peaks in criminal activity are observed during the mid-to-late Kangxi reign and the late Qianlong period, with high case density in populous regions such as Shandong, Jiangnan, and Guangdong. This study employs digital humanities approaches to analyze unstructured historical archives, exploring the potential of applying computational methods to the semantic interpretation of criminal motives and the broader field of judicial history. The research aims to provide alternative methodological perspectives for engaging with historical sources.

Keywords

Digital humanities, Qing Dynasty Yellow Registers, Household registers, Criminal records, Text mining

【 Summary 】

Introduction

The Qing Dynasty (1644-1912), as one of the longest-lasting imperial regimes in Chinese history, developed a highly institutionalized legal and administrative system that reflected the social order, moral norms, and governance strategies of its time. Among the diverse archival materials preserved from this period, the *Yellow Registers* (huangce) constitute a particularly valuable yet underexplored source for understanding state administration and social dynamics. Originally developed during the Ming dynasty as household and taxation registers, the Yellow Registers in the Qing period expanded significantly in scope, encompassing a wide range of administrative records, including criminal case reports submitted to the central government.

Despite their richness, prior scholarship on Yellow Registers has largely focused on fiscal administration, household registration, or institutional structures, with relatively limited attention paid to their criminal case records. Moreover, traditional historical studies have tended to rely on close reading of selected cases or institutional analyses, making it difficult to grasp broader patterns across large corpora of judicial documents. In this context, the emergence of digital humanities (DH) methodologies offers new opportunities to analyze large-scale historical datasets systematically.

This study adopts a digital humanities approach to examine criminal case records embedded within Qing Dynasty Yellow Registers. By integrating generative artificial intelligence (GAI), Python-based text analysis, and the DocuSky platform, the research aims to explore three key dimensions: (1) the typology of crimes, (2) the semantic patterns of criminal motives, and (3) the spatiotemporal distribution of criminal cases. Through this approach, the study seeks to demonstrate how computational methods can transform unstructured historical archives into analyzable data and provide new perspectives for judicial and social history research.

Methods

The dataset used in this study consists of 1,745 Chinese-language Yellow Register entries preserved in the digital archives of the Institute of Modern History, Academia Sinica. After removing duplicate yearly entries, a total of 1,698 valid records were retained, covering the period from 1664 (Kangxi 3) to 1820 (Jiaqing 25). The analysis focuses on three key metadata fields: case summaries, temporal information (reign years), and administrative jurisdictions.

The research workflow involves three main stages: data preprocessing, semantic classification, and spatiotemporal analysis.

First, data preprocessing was conducted using Python. The original records were cleaned, normalized, and converted into a structured CSV format, with each case represented as a single entry. Text segmentation and keyword extraction were applied to facilitate further analysis.

Second, crime types and motives were identified through a hybrid approach combining GAI-assisted semantic interpretation and rule-based classification. The ChatGPT 4-o model was employed to interpret case summaries and extract relevant action verbs and contextual cues, enabling the classification of crimes into categories such as homicide, assault, sexual offenses, theft, and robbery. In parallel, a custom Python dictionary was constructed to perform keyword matching and frequency analysis. Crime motives were similarly categorized into themes such as domestic disputes, romantic conflicts, financial disputes, and retaliatory actions. Batch processing (100 cases per iteration) was used to ensure consistency and scalability.

Third, temporal and spatial analyses were conducted. Temporal patterns were examined by aggregating cases according to reign periods and years. For spatial analysis, administrative locations mentioned in the records were standardized and geocoded. Using DocuSky's DocuGIS module, these locations were mapped onto Qing-era administrative divisions and visualized to reveal regional distributions and clustering patterns.

It should be noted that this study is exploratory in nature. Due to limitations

in data completeness, linguistic ambiguity in classical Chinese, and the absence of large-scale manual validation, the results are interpreted as indicative patterns rather than definitive historical conclusions.

In addition, this study proposes a hybrid analytical framework that combines rule-based keyword classification with GAI-assisted semantic interpretation. This approach allows for both scalability and contextual sensitivity, addressing the limitations of purely dictionary-based methods when applied to classical Chinese texts. By integrating computational efficiency with semantic flexibility, the framework provides a replicable model for analyzing large-scale historical archives in digital humanities research.

Results

The analysis reveals several notable patterns in the crime records of Qing Dynasty Yellow Registers.

First, in terms of crime typology, violent crimes dominate the dataset. Homicide and assault constitute the most frequent categories, significantly outnumbering property-related crimes such as theft and robbery. This suggests that the Yellow Registers, particularly those submitted for imperial review, tend to emphasize serious offenses involving bodily harm or death, reflecting both administrative priorities and reporting practices.

Second, regarding criminal motives, three primary categories emerge: domestic disputes, romantic conflicts, and spontaneous altercations. Domestic disputes often involve family tensions, including conflicts between spouses or relatives, while romantic conflicts frequently relate to extramarital relationships or disputes over marriage arrangements. Spontaneous altercations typically arise from verbal arguments or minor confrontations escalating into violence. These findings highlight the centrality of interpersonal relationships and emotional dynamics in Qing criminal cases, suggesting that many violent incidents were embedded within everyday social interactions rather than organized criminal activities.

Third, temporal analysis indicates that criminal case records exhibit noticeable fluctuations across different reign periods. Peaks are observed during the mid-to-late Kangxi reign and the late Qianlong period. These fluctuations may reflect broader socio-political conditions, population growth, or administrative reporting practices, although further historical contextualization is required to fully interpret these trends.

Fourth, spatial distribution analysis shows a clear concentration of cases in densely populated and economically developed regions, including Shandong, Jiangnan, and Guangdong. These regions not only had larger populations but also more active administrative reporting systems, which may contribute to their higher representation in the dataset. The geographic clustering of cases suggests that both demographic density and administrative infrastructure play important roles in shaping the archival record of crime.

Overall, the results demonstrate that combining semantic analysis with quantitative and spatial visualization techniques can effectively uncover macro-level patterns in historical judicial data.

Conclusion

This study illustrates the potential of digital humanities methodologies in transforming and analyzing large-scale historical archives. By integrating generative AI, Python-based text mining, and geospatial visualization tools, the research provides a systematic exploration of crime typologies, motives, and spatiotemporal patterns in Qing Dynasty Yellow Registers.

Methodologically, the study demonstrates that GAI-assisted semantic classification can complement traditional keyword-based approaches, enabling more nuanced interpretation of historical texts, particularly those written in classical Chinese. At the same time, the use of computational tools allows for scalable analysis, making it possible to move beyond case-based studies toward broader, data-driven observations.

Substantively, the findings suggest that violent crimes rooted in

interpersonal conflicts were a dominant feature of Qing criminal cases, and that these incidents were shaped by both social relationships and regional contexts. The observed temporal and spatial concentrations further highlight the interplay between historical conditions and archival representation.

However, the study also underscores several limitations. The reliance on summarized metadata rather than full original texts may affect interpretive depth, while the absence of systematic accuracy evaluation for AI-based classification introduces potential uncertainty. Moreover, the dataset represents only a subset of surviving archives, limiting the generalizability of the findings.

Despite these constraints, this research contributes to the growing field of digital humanities by offering an exploratory framework for analyzing judicial archives. It demonstrates how computational approaches can generate new insights into historical crime patterns and provides a foundation for future studies that may incorporate larger datasets, more rigorous validation methods, and deeper historical interpretation.

More broadly, this study highlights the role of digital humanities as a bridge between computational methods and historical interpretation, suggesting that the integration of AI technologies may reshape how historians engage with large-scale archival materials.

壹、緒論

清代作為中國歷史上統治時間最長的朝代之一，其司法制度與刑事案件的運作深刻地反映了當時社會的秩序與人倫關係。特別是黃冊檔案中所保留的刑事案件紀錄，不僅詳細記載了罪犯、案件發生地、作案手段與社會背景，亦蘊含了清代社會治理模式與地方社會運作的珍貴史料價值。黃冊是始於明代的一種戶籍和賦役檔案，主要用來記錄人口、田產和賦役情況，因其以黃色封面冊子裝訂，故名黃冊。清代沿襲舊制，「按清制，凡屬進呈御覽之冊，冊面均為黃色，統稱黃冊，其應用範圍較明代更廣，內而部院衙門，外而各省臣工，不限戶籍、賦役，凡關乎報銷錢糧，或例應造報事件，均須造具黃冊，隨同題本進呈御覽」（趙郁楠，2018）。而清代黃冊的範圍極廣，並不限於戶籍、財政與行政事務，其中亦包含刑部所呈之重案、死刑案件。本文所使用的「清代黃冊檔案」即屬於此一龐大體系中關涉刑案之部分，透過數位人文方法的處理，可以觀察清代刑事案件的類型、動機與時空分布，在過往研究關注經濟、戶籍資料之餘，期許能提供對於清代刑案的觀察。

過往學術界對於黃冊檔案內容的研究相當有限，在為數不多的論文發表主題中，有探討婦女戶籍登記、軍戶家丁、里甲制度等（李新峰，2010；于志嘉，2012，2018；衣若蘭，2024），多半是集中於特定主題的文獻探討和內容介紹，尚且缺乏以量化數據方式呈現黃冊檔案整體內容概況。而近年崛起的數位人文研究，適可提供對於史料檔案進行綜觀分析的不同研究取徑。研究者可運用數位工具輔助，進行以往耙梳大量文本難以達成的研究，例如可運用主題模型、社會網絡分析（**Social Network Analysis**）以及地理資訊系統等工具，從檔案中發掘隱含主題、還原歷史人物的社會互動，或再現事件的時空脈絡，或可開啟不同的研究分析視角與知識傳播的模式。

目前在臺灣留存的清代漢文黃冊檔案，主要典藏在中央研究院歷史語言研究所，本研究以中央研究院近代史研究所的清代奏摺檔案資料庫中的漢文黃冊檔案為分析對象，檔案內容為地方各省犯罪登記事件，本研究進行初步清理與斷詞處理後，進行犯罪行為詞彙、地名的辨識，並結合 Python、GAI 與 DocuSky 等數位工具進行統計與視覺化分析，以全面瞭解

清代各地、各年代的犯罪案件分布趨勢及特徵，並進一步探討社會犯罪動機的多樣性及其時代脈絡。

從清代犯罪史研究的角度而言，過往研究多依賴個案式史料分析或制度史研究，較少以大量司法文書資料進行整體量化觀察。本研究意義在於實驗性探討此批資料的數位化運用，意在提供對於應用數位典藏的檔案內容進行鳥瞰分析的研究案例，而非對於歷史事件的專深解釋，對於清代刑案的研究，尚需內容專家根據文本與史實，進行更深度的研究解釋。本研目的在於：

一、實驗性運用 GAI 模型與 Python 進行犯罪語義分類與犯罪動機識別分析

探索 GAI 語意分類模型在歷史語境下的適用性，進行犯罪類型與動機的初步分類，評估其於清代黃冊文本中辨識隱含語義與社會行為模式的潛力。

二、運用 Python 與數位人文工具，分析清代黃冊中的犯罪時空分布與地理趨勢

透過地理與時間資訊的視覺化，呈現不同區域與時期的犯罪熱點與變化趨勢，理解歷史檔案中的犯罪活動與空間、時間因素的交織。

三、探討數位人文研究方法在歷史檔案知識生成中的應用潛能

透過黃冊檔案的數位處理與分析流程，對黃冊刑案進行初步量化分析，盼驗證數位工具如何協助轉化非結構化歷史文本為可計算、可詮釋的資料資源，拓展人文研究的方法論，並嘗試辨識官方檔案敘事中呈現的犯罪類型與動機模式，為清代犯罪史研究提供新的觀察尺度。

貳、文獻回顧

隨著數位技術的發展，歷史文本的整理與詮釋方式也出現重大轉變。近年來，歐美與臺灣學界紛紛透過數位典藏、資料探勘、地理資訊系統與

視覺化等技術，發展出多樣化的歷史文本分析方法，不僅提升研究效率，也開啟了跨領域合作與史學創新的新局。本文之文獻回顧將以三層面分述，首先聚焦研究對象，回顧以清代黃冊檔案之相關文獻；其次，梳理國內外運用數位人文技術分析歷史犯罪或司法檔案的先驅性案例；最後，綜覽數位人文在廣義的歷史檔案研究中之多元應用。

一、清代黃冊與相關研究

黃冊又稱賦役黃冊，是明洪武十四年（1381年）開創的戶籍與賦役之法的制度，因套用黃色封面，故總稱「黃冊」。黃冊以戶為單位，詳細登載籍貫、姓名、年齡、丁口、田宅、資產等戶籍資料。黃冊照式樣編造四份，分別由戶部、布政司、府、縣各保存一份。保存在戶部的一份，冊面是用黃紙，故而通稱為黃冊。黃冊，是明朝戶籍制度的核心部分，對於國家徵收賦稅和徭役具有重要作用（趙郁楠，2018）。

清朝定滿語為國語，從中央到地方的各級滿蒙官員，特別是承辦八旗事務及邊疆少數民族事務的滿蒙官員，一般都用滿文繕寫公文，內閣作為輔佐皇帝辦理國家政務的中樞機關，每天承接辦理的各種公文中有大量的滿文文書，故而形成了為數不少的滿文黃冊。按清制，凡屬進呈御覽之冊，冊面均為黃色，統稱黃冊，其應用範圍比明代更廣，內而部院衙門，外而各省臣工，不限戶籍、賦役，凡關乎報銷錢糧，或例應造報事件，均須造具黃冊，隨同題本進呈御覽。因此，清代內閣滿文黃冊是作為題本附件而進呈御覽的官文書檔冊（趙郁楠，2018；劉錚雲，1998）。

清代內閣黃冊中，涉及京城皇族、在京八旗文武職官兵及邊疆少數民族類事務，多用滿文繕寫；涉及各省人民戶口、賦役事務，綠營與駐防八旗兵馬、錢糧事務，多用漢文繕寫。究其因主要是在京各部院繕寫呈進的題本稱為「部本」，是為滿文，而各省將軍、總督、巡撫等員繕寫呈進的題本稱為「通本」，多為漢文。於今觀之，漢文題本內容更為豐富，多為各省督撫提報民丁屯丁戶口、地丁錢糧、漕費鹽課及綠營兵馬、錢糧、勘合數目等。在清代，黃冊的應用範圍極其廣泛，可謂適用於京內外各衙門題奏事項之附冊，凡河工報銷、營建工程、錢糧報銷、職官京察、朝審秋審等均在其中，具有非常重要的史料價值（于志嘉，2018；趙郁楠，2018）。

根據國立北平故宮博物院於 1936 年所編《內閣大庫現存清代漢文黃冊目錄》，現存清內閣大庫漢文黃冊六千餘本，均為京內外各衙署隨本進呈，經御覽後交出存閣。其內容以報銷收支款項為最多，大抵分為「舊管」、「新收」、「開除」、「實在」四項，稱為「四柱」；戶部三庫另有「大進」「大出」兩類。此外，亦有事不關會計，而屬行政業務照例繕呈之黃冊，但數量相對較少。清制規定，凡全國庶政例應奏報者，主管衙署均須依限期繕造黃冊進呈御覽，另具青冊副本分送部科以資審核。黃冊所載雖多為簡要數據，但編排方式已漸具近代統計雛形，足以觀察清代法政、經濟制度的沿革全貌。該館整理時以衙署為單位，依各冊性質分類列目，按年排比，計有吏、戶、禮、兵、刑、工六曹及都察院、理藩院、太常寺、光祿寺等，合計五十一類、一百二十四目（國立北平故宮博物院文獻館，1936）。而國立北平故宮博物院文獻館、國立北京大學文科研究所、國立中央研究院歷史語言研究所合編之《清內閣舊藏漢文黃冊聯合目錄》中，則載有自順治迄光緒各朝，凡一萬七千零三十三冊，七十五類，一百六十四目（國立北平故宮博物院文獻館等，1947）。此一聯合目錄為紙本，在臺灣提供清內閣漢文黃冊目錄查詢之數位典藏系統，為中央研究院近代史研究所（簡稱中研院近史所）數位資料庫，本研究採用之資料即為中研院近史所資料庫中所有的黃冊檔案，共 1,745 筆。

二、數位人文運用於刑事犯罪檔案研究

隨著數位技術在歷史學與人文研究中的應用日益成熟，運用歷史刑事檔案與法律文書進行數位分析，逐漸成為犯罪史與司法史研究的重要趨勢。相關研究透過文本數位化、資料庫建置、資料探勘與視覺化工具等方法，嘗試從大量司法紀錄中重建犯罪類型、司法制度與社會控制機制的歷史樣貌。

Mao 與 Ma (2012) 則從數位化檔案整合與跨機構協作角度，分析《新清史》計畫下清代檔案數位化工程的發展。該計畫整合中國大陸三十餘類檔案庫與逾百萬筆文件，建立多欄位檢索系統，使研究者得以跨區域串聯地方司法檔案與中央奏摺資料。研究指出，數位化技術大幅提升宏觀歷史分析的可能性，但亦揭示檔案全文檢索不足與資料存取權限限制等問題，顯示清代檔案數位化仍面臨制度與技術層面的挑戰。

Park (2020) 以法律史學、文本分析與比較研究方法，對清代官員法律體系中的「公罪」與「私罪」概念進行系統性分析。該研究主要透過《大清律例》、《大清會典》及六部則例等法典與行政規章之細緻解讀，並輔以帝王諭旨與官員奏議等檔案文獻，探討清代如何透過普通刑法與官僚專用法規並行的「雙軌制法律框架」，界定與處置官員違法行為。研究指出，「公罪」與「私罪」的區分涉及情境、動機與後果等判定要素，並進一步分析其在司法程序與量刑制度上的差異。透過縱向梳理周代至明清法律觀念之演變，以及橫向比較歐洲行政法體系，該研究指出清代法律制度最終服務於維繫官僚效能與王朝治理穩定的政治邏輯，為理解清代司法行政運作提供制度的脈絡。

除清代檔案研究外，西方犯罪史領域亦發展出成熟的數位研究典範。Guiney (2020) 以英格蘭與威爾斯假釋制度為例，透過政府公文、檔案館藏與議會紀錄等多元資料來源，分析刑罰政策轉變與制度演化。研究採用默頓 (Robert K. Merton) 的中層理論 (Theories of the Middle Range)，透過具體歷史案例連結政策制度與行政決策過程，並指出數位轉型使政府紀錄呈現碎片化現象，同時也促進質性研究電腦輔助軟體 (Computer-assisted qualitative data analysis software, CAQDAS) 與社會網絡分析等工具在犯罪學研究中的應用。

「追蹤倫敦罪犯：十八至二十世紀英澳刑事軌跡」(Digital Panopticon: Tracing London Convicts in Britain & Australia, 1780-1925) 計畫則建置跨國罪犯資料庫，整合英國與澳洲檔案館資料，透過 OCR 與人工校對確保文本辨識率，並提供多維度查詢與視覺化分析功能，為犯罪遷徙與刑罰制度研究提供重要資料基礎。(Digital Panopticon, 2020) 此外，「舊貝利刑事法庭審判記錄線上資料庫」(The Digital Humanities Institute of University of Sheffield, 2023) 計畫，系統性數位化倫敦中央刑事法院 1674 至 1913 年間審判紀錄，並透過 XML 標記技術建立結構化資料庫，支援犯罪類型、判決結果與人物關係等量化研究，成為數位犯罪史研究的重要基礎資源。而在華人地區司法檔案研究方面，《淡新檔案》數位人文研究亦展現重要成果。胡其瑞 (2024) 運用 THDL 與 DocuSky，對清末宗教與司法案件進行文本標記與 GIS 空間分析，揭示教會傳播網絡與地方社會互動，並討論數位文本分析中名稱變體造成之資料連結問題，突顯權威控制與詮釋方法的

重要性。

整體而言，刑事與司法檔案之數位人文研究，已從資料庫建置與文本數位化，進一步發展至制度分析、社會網絡重建與空間視覺化分析，顯示數位方法在犯罪史與司法史研究中的高度潛力。然而，相較於西方犯罪史資料庫之成熟發展，清代司法檔案之數位量化研究仍相對有限，亦凸顯本研究以數位方法進行清代司法檔案內容分析具有史料探索上之學術意義。

三、數位人文運用於文史檔案研究

除刑事司法檔案外，數位人文方法亦廣泛應用於各類歷史文本與文史檔案研究，透過自然語言處理、資料探勘與社會網絡分析等技術，協助研究者從大量非結構化文本中萃取人物關係、知識傳播與文化網絡等歷史現象，為傳統史學研究提供新的分析視角。

Hsiang 等人 (2012) 以「臺灣歷史數位圖書館」(THDL) 收錄之 37,831 件清代檔案為研究對象，透過創新演算法挖掘檔案文本中的制度運作關係。研究首先利用奏摺與諭旨之格式標記，自動辨識文件引用關係，成功建立 6,801 組文件配對，並進一步建構 1,101 個「諭旨－奏摺關係圖」，呈現清代中央與地方行政決策互動過程。此外，研究透過格式相似性分析，辨識出官僚文書中高度標準化的行政模板，並將模板差異資訊轉換為可量化資料，用以分析歷史事件之時序變化與制度回應，例如林爽文事件期間米價變動與捐監生制度調整。該研究不僅突破清代檔案因分散典藏與編年體歸檔造成的脈絡破碎問題，也展示數位方法在制度史與事件史研究中的應用潛力。

關伯丞等人 (2012) 以《清聖祖實錄》為研究對象，運用資料探勘技術自動生成職官年表，並進一步分析歷史人物社群網絡。研究透過頻繁區間項目集合探勘演算法，自動識別職官名稱與任期資料，並提出循序共陸樣式探勘方法，重建歷史人物仕途關係。該研究展示資料探勘技術在歷史文獻分析中的應用潛力，並顯示跨領域方法在史學研究中的重要性。許雅惠 (2018) 則透過 CBDB 與社會網絡分析工具，重建北宋金石收藏圈之人際網絡，並分析士大夫與官方收藏權力轉移過程，進一步展現數位方法在藝術史與文化史研究中的應用潛力。

同樣在人物網絡研究方面，陳冠至等人 (2018) 運用「中國歷代人物

傳記資料庫」(CBDB)為基礎,運用社會網絡分析探討明代蘇州藏書家社群之發展與變遷。研究透過網絡密度與中心性分析,呈現社群結構與知識傳播模式,顯示數位人文方法在文化史與知識史研究中的應用價值。後續陳柏溢與陳冠至(2021)亦以CBDB資料,透過社會網絡分析方法探討清代紹興藏書家群體之社會關係結構,指出藏書家之間透過血親、姻親、師承與交遊關係形成文化網絡,並藉由中心性指標分析關鍵人物角色,呈現地域性知識社群之形成過程。

在思想史研究方面,邱偉雲(2020)運用自然語言處理技術分析「中國近現代思想史專業資料庫」,透過高頻共現詞分析與時序語義變化,探討近代「天道」概念之思想轉型,展示量化文本分析在觀念史研究中的應用可能。

而社會經濟史研究方面,趙思淵與潘芸淇(2024)以明清契約文書為研究材料,運用DocuSky平台進行全文標記與地理資訊分析,揭示土地交易與社會信用體系之運作,並展示數位環境建置對史料觀察方式的改變。

整體而言,文史檔案數位人文研究展現出強大的跨領域整合能力,無論是在人物網絡重建、思想史語義分析,或契約文書與文化網絡研究上,均顯示數位工具可有效提升史料分析效率,並拓展歷史詮釋之可能性。本研究亦在此基礎上,嘗試將相關數位方法應用於清代刑事檔案分析,以探索犯罪類型與犯罪動機之語義結構與時空特徵。

參、研究設計與實施

一、研究對象

本研究所採用之資料,主要來自中央研究院近代史數位資料庫中的「清代奏摺檔案資料庫」。該資料庫收錄自清代官方奏摺、黃冊等檔案,可區分為「財稅經濟」與「刑事案件紀錄」兩大類。財稅經濟資料包含稅收、戶籍、糧餉等經濟活動之相關文書;刑事案件紀錄則包含刑科題本及黃冊檔案,具體反映清代各地社會治安與法律制度的實際運作(中央研究院近代史研究所, n.d.)。透過系統化整理與分析,可進一步窺探清代社會的刑事犯罪樣態及其時代特徵。

本研究選取之標的文本「漢文黃冊檔案」，係由刑部審理案件後編纂成冊，詳載案件背景、罪犯籍貫與具體罪行，且經數位化後收錄於中研院近代史數位資料庫者，共計 1,745 件。經查部分案件可能因審理或偵查進度，存在於不同年度重複紀錄之情況，此處則僅保留第一次出現之年份資料，共計 1,698 筆。案件分布年代起自康熙三年（1664），終於嘉慶二十五年（1820），包括「刑部重囚招冊」、「刑部秋審犯人招冊」、「刑部各省重囚招冊」、「刑部秋審福建省情真犯人招冊」等 137 冊，為研究清代刑事司法實務與社會狀況之重要參考。該資料庫原始記錄包含 10 個欄位，分別為案件主題、案件摘要、朝代／年代、司名、罪犯、籍貫、卷名、卷號、冊號、冊主題及資料庫名稱。本研究以案件摘要、朝代／年代、司名三個欄位為分析標的，此三個欄位內容載錄案件背景與罪行，足可供案發地點、年代與案件類型之探索。

二、研究工具

本研究結合多種數位人文研究工具與自然語言處理技術，進行清代黃冊刑案資料的文本分析與視覺化處理。研究工具包括 Python、生成式人工智慧 Chat GPT 4-o 模型以及 DocuSky。各工具之功能與本研究之應用如下說明：

（一）Python

Python 為本研究中資料處理的主要工具。其廣泛的文字處理函式庫（如 pandas、re、collections）與統計套件，使其特別適用於大型文本資料的清理、轉換與分析。首先以 Python 將黃冊資料進行預處理，再依 metadata 欄位結構轉換為 CSV 格式，並進行資料清理。其次，Python 亦用於自建犯罪類型與動機的分類詞典，並透過字詞比對與正規表達式進行文本自動標註與詞頻統計。此外，Python 亦用於時間與地理欄位資料整理統計，進行跨年代與跨區域的案件分布分析，奠定時空間分析的基礎。

（二）ChatGPT

為進一步提升文本理解與分類的深度，本研究亦採用 OpenAI 所開發

之生成式人工智慧 ChatGPT 4-o 模型。ChatGPT 在自然語言理解與上下文推理方面具備高度準確性與語義感知能力，能夠協助針對案件摘要中的語句進行語意解讀與犯罪意圖判別。研究中，ChatGPT 4-o 應用於輔助閱讀摘要內容，並根據語義與情境將案件細分為不同犯罪類型與動機類別，補足人工詞典難以涵蓋的語境變異，強化分類的靈活性與語意敏感度。

（三）DocuSky

DocuSky 為國立臺灣大學數位人文研究中心、資訊工程學系數位典藏與自動推論實驗室所開發之文本分析與資料視覺化平台，本研究利用 DocuSky 的地理資訊處理功能 DocuGIS，將文本中經過實體辨識、人工審核與標準化的地名實體對應至清代行政區劃，進一步生成地理資訊分布圖。藉由空間標示呈現犯罪案件的分布熱點與地理變化趨勢，有助於分析犯罪現象之空間特性及其歷史脈絡。

三、資料分析過程

（一）資料預處理

本研究首先透過 Python，將「清代奏摺檔案資料庫」中的黃冊檔案資料進行萃取與清理，保留案件摘要、朝代／年代、司名三欄位，並轉換為一行一筆的 CSV 格式，作為後續分析之依據。

（二）犯罪類型與動機統計

完成文本預處理後，本研究將案件初步劃分為殺人、傷害、性犯罪、盜竊、搶劫等類別。犯罪類型部分，先以一次 100 筆為單位，由 GAI 分批閱讀，進行犯罪相關動作詞彙（如打死、刺死、鬥毆、偷竊等）之提取與分類，再進一步依據所提取詞彙內容加以細分。接著，透過 Python 建立分類詞典，計算各類犯罪相關詞彙之詞頻統計，藉此觀察各類型犯罪的分布情況。犯罪動機部分，則依據案件摘要所述之犯罪動機關鍵詞，初步劃分為家庭糾紛、感情糾紛、財務糾紛、挾怨報復等類別，同樣以一次 100 筆為單位交由 GAI 閱讀，進行犯罪動機之整理與統計。

(三) 年代與地理資訊統計

先針對資料中的朝代／年代欄位進行統計，以觀察各年代的犯罪分布趨勢。同時將審理司名欄位剔除重複詞，建置為詞典，進行詞頻統計，於 Google Earth 人工檢索各司對應之現今各省經緯度後，結合 DocuSky 中 DocuGIS 功能，繪製成地理資訊分布圖，呈現各地區犯罪案件的空間分布情形。此外，將各年度分隔後，運用 Python 統計各年度與審理司名資訊，亦可統計出各年度下不同地區之犯罪情況，輔助進行跨時空分析。

整體而言，本研究重點在於資料與方法的探索，採用數位人文方法，包括文本分類、視覺化及時空分析，以揭示漢文黃冊中案件的分布與模式，且可能囿於臺灣目前典藏的數位檔案數量限制，無法盡觀完整的黃冊檔案內容，因此，對於分析結果的解釋，僅限於觀察臺灣典藏檔案內容所呈現的樣貌，不代表完整歷史因果分析或事件重建的解釋。

四、研究限制

本研究透過數位人文方法系統性分析清代黃冊檔案，然因史料特性與技術限制，仍存在以下限制：

(一) 犯罪動機與類型分類依關鍵詞與 GAI 輔助判讀，然對清代文書的句構、詞彙習慣、語用特徵乃至筆誤、簡寫者，未必能完全準確辨識，可能對結果產生細微影響。例如部分案件僅記載「因口角毆斃」，缺乏具體衝突背景，使模型難以判斷其屬於家庭糾紛或偶發衝突，顯示古漢語司法敘事的簡略性亦可能影響語義分類結果。

(二) 本研究並未進行大規模人工逐筆標註與模型分類結果之系統性量化比對(如一致率或召回率分析)，且受限於資料完整性，如檔案不連續、缺年、同一人名於不同歲次重複呈列等，因此不以統計方式精確評估模型之分類準確度。結論應視為基於現存樣本分布的探索性結果，而非對清代黃冊刑事現象的絕對描述。

(三) 本研究所使用之「摘要」欄位，多為中研院近史所典藏與整理時所加之後設資訊，而非清代官方原有之正式類目。其資料整理方式可能反映早年研究人員之邏輯，未必能完整呈現原始黃冊之編纂體例。因此，本研究之統計僅為對現存資料樣本的嘗試性呈現，而非對清代中央刑部呈報制度之全貌重建。

肆、研究結果

一、犯罪類型分布

運用 GAI 對黃冊文本進行語義分類後，以 Python 進行詞頻統計，結果顯示殺人與傷害類暴力犯罪為主要類型，其次為性犯罪與財產犯罪（表 1）。殺人相關詞彙如「打死」（212 次）、「致死」（155 次）、「戳死」（141 次）、「謀殺」（107 次）等出現頻率最高，反映清代社會中暴力致死事件的多樣性與嚴重性。「傷害」詞彙如「毆傷」（235 次）、「戳傷」（111 次）、「致傷」（113 次）等亦顯著，顯示非致命性暴力亦為常見案件型態。性犯罪方面，「通姦」（122 次）居首，其次如「雞姦」（46 次）、「輪姦」（10 次），呈現以婚姻倫理與性暴力為主之紀錄特徵。財產犯罪詞頻以「偷竊」（110 次）、「行竊」（67 次）等竊盜情形較高，但部分如「行劫」（35 次）亦顯示與暴力結合之情形。其他類型如「拒捕」（35 次）、「誣告」（11 次）則是呈現清代亦有司法對抗行為（拒捕）與行政資源濫用（誣告）的情形。

表 1

GAI 分類之各類犯罪關鍵詞暨詞頻統計

案件類別	詞彙	詞頻	案件類別	詞彙	詞頻
殺人	打死	212	性犯罪	通姦	122
	戳死	141		輪姦	10
	致死	155		雞姦	46
	謀殺	107		調姦	4
	謀死	70	盜竊、搶劫	偷竊	110
	砍死	81		行竊	67
	勒死	75		竊盜	14
	扎死	35		偷盜	19
	刺死	25		搶奪	17
	踢死	10		盜犯	8
	毒死	11		竊賊	14
	淹斃	11		強賊	8
	溺死	7		行劫	35

（續下表）

（接上表）

案件類別	詞彙	詞頻	案件類別	詞彙	詞頻
傷害	毆傷	235	其他罪行	拒捕	35
	戳傷	111		誣告	11
	砍傷	65			
	扎傷	58			
	致傷	113			
	踢傷	16			
	誤傷	15			
	打傷	71			
	共毆	27			
	鬥毆	46			

二、犯罪動機分布

本研究嘗試以 GAI 針對黃冊檔案中所載之殺人案件進行動機分析，透過案件文本內容的語言線索與敘事脈絡，重構行為動機之分類架構。為輔助初步辨識，透過指令與案例試閱（約 500 筆）讓 GAI 逐步熟悉案件摘要之呈現內容，建立語言線索對照表，配合人工預先建立之各動機類型（如財務糾紛、感情糾紛、家庭糾紛、工作或主僕衝突、積怨報復、偶發衝突、精神異常／無明動機，以及教唆殺人／集體鬥毆），設計出對應典型詞彙與句式結構，以利動機判讀，以 100 筆案件為單位分別匯入，獲取犯案動機整理，最終進行加總統計。惟本研究分析為數位人文研究範疇中之實驗性探索，量化呈現犯罪類型與動機分布應理解為對「官方檔案敘事樣態」之量化觀察，而非對於歷史行為者真實心理動機之直接還原，若欲以史實重建為核心之歷史論述，尚有待與歷史學者合作針對特定議題補強論述深度。

本研究分類過程中對於語意模糊或動機重疊之案例，需依具體情節推斷其主要衝突來源。但因犯罪行為樣態複雜，許多案件動機為疊加形成，如家庭內糾紛犯案，可能同時存在情感糾紛或財產糾紛等。因此，多重動機之案件，本研究採重複計入不同動機。例如「張行已毆死伊妻王氏一案」，導火線為嫌犯張行已向受害人王氏（其妻）討取衣物首飾典當，王氏反對

並斥責丈夫。而張行已惱羞成怒，持棍棒、磚頭將王氏打死，還將銀鏞灌入王氏口中，試圖假冒輕生。此案動機便同時包括家庭糾紛與財務糾紛，正可體現分類方法中語義、脈絡與邏輯判讀三者交織運作之應用實例。本研究分析動機類型與判別的線索說明如下（表 2）。

表 2

犯罪動機類型與判別的線索說明表

動機類型	說明	判別線索
財務糾紛	因金錢、財物爭奪或債務糾葛引發的殺人行為	欠銀、討債、借錢、銀兩、錢不還、貨不交、錢財糾葛、贖當、交易失信、分贓爭執
感情糾紛	涉及戀情、配偶爭奪、姦情或婚配衝突等情感因素	姦、私情、舊情、婚配、爭妻、姦宿、戀愛糾紛、私約、配婚不滿、通姦被撞見
工作／主僕衝突	服役、勞務或奴僕與主家間之權責及待遇問題所致之爭執	主子、奴僕、僕人、使喚、不配、解職、罰役、差遣、役使不公、虐僕、解庄
家庭糾紛	父子、夫妻、兄弟、妯娌及姻親間的衝突	父子、兄弟、夫婦、妯娌、姻親、家中口角、家人不合、子殺父、夫殺妻、兄弟成仇
積怨報復	長期積怨、仇殺或報仇行為	懷恨、成仇、報仇、仇殺、舊怨、宿怨、夙怨、曾打鬥、曾控告、相互爭訟已久
權利糾紛 （地位／物件）	家產、地位、妻妾或僕從等歸屬的爭奪	爭地位、家產不均、爭僕人、分家糾葛、爭妻妾、奪產、分配不均、主家之爭
偶發衝突 （口角／酒後 ／即興殺人）	無重大背景的臨時爭執	吃酒、口角、酒後、即興、鬥嘴、撞見、臨時爭執、打鬧、無因動手、偶遇、當場爆發
無明確動機／ 精神異常	動機含混、模糊，或疑似精神異常	動機不明、胡亂打人、神志不清、無原因、自語、自行承認不知何故下手、無前因
教唆殺人／ 集體鬥毆	由他人煽動或多方人馬混戰中之殺人事件	領人、聚眾、煽動、十餘人、群毆、商議要殺、串連行動、受鼓動、勾結數人、會眾鬥毆

而視覺化結果（圖 1）顯示，「家庭糾紛」、「感情糾紛」、「權利糾紛」與「偶發衝突」為最主要動機類別，佔總案件數逾半，其中又以家庭内部的衝突為殺人事件最常見之根源，可能呈現家庭矛盾與經濟壓力交織的特徵。而其次的感情糾紛與偶發衝突，則顯示私人情感與衝動型暴力在案件中具有高度影響力。性質具相似性的「權利糾紛」與「財務糾紛」同樣明顯地占有較高比例。「工作／主僕衝突」等，則往往與主要動機交錯出現，反映清代社會中可能存在之人身隸屬、資產歸屬與社會階序之張力。整體而言，案件動機呈現複合性與連動性，需結合主次分類進行交叉分析，以更全面理解犯罪行為生成脈絡。

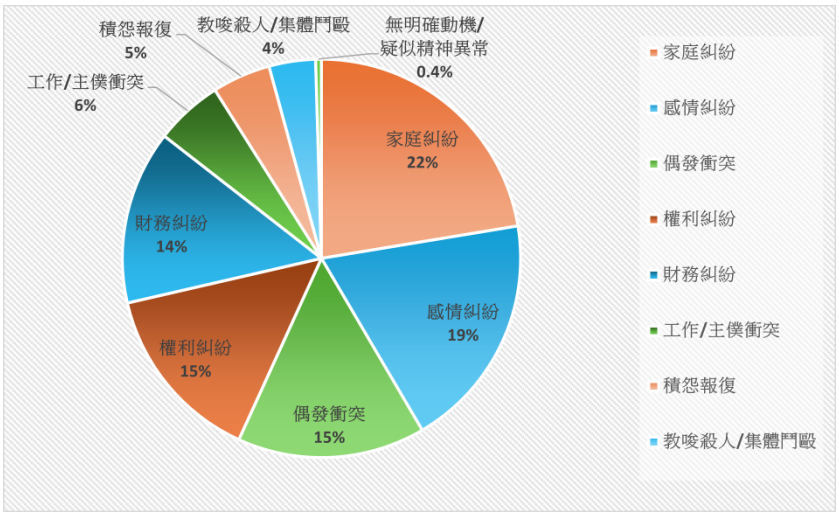


圖 1 犯罪動機分布圖（含同一案件之重複統計）

三、各年度案件總數分布

黃冊案件數量的分布呈現階段性波動（圖 2），與當時的社會環境、治安情勢及司法管理等因素相關。而經統計，康熙年間案件數量呈現較明顯波動，其中康熙 20 年（1681 年）達到高峰（90 件），康熙 30 年（1691 年）、康熙 31 年（1692 年）及康熙 61 年（1722 年）也分別達到 50 件以上，反映該時期社會衝突較多或司法機構運作較為活絡。雍正年間（1723-1735

年)案件數量較康熙略低,多數年份維持在 10 至 40 件間,相較於其他時期,雍正年間犯罪案件數相對較少,可能與雍正帝整飭吏治、強化司法管理有關。乾隆年間(1736-1795 年)初期則呈現相對穩定的態勢,如乾隆 3 年(1738 年)、乾隆 10 年(1745 年)皆達 50 件以上,中後期漸有下降。嘉慶年間初期(1796-1820 年)案件數量較少,但嘉慶 24 年(1819 年)及嘉慶 25 年(1820 年)各有 50 件,似乎可見嘉慶後期的案件高於初期的案件審理數。需特別指出,黃冊所記錄之案件數量反映的是「官方上報並存檔之案件數」,而非社會中實際犯罪發生率。案件數量或來源地點亦可能受各地方行政體系、文書保存與司法制度運作等因素影響,故應理解為對官方檔案分布情形的觀察,而非對社會實況之直接測量。

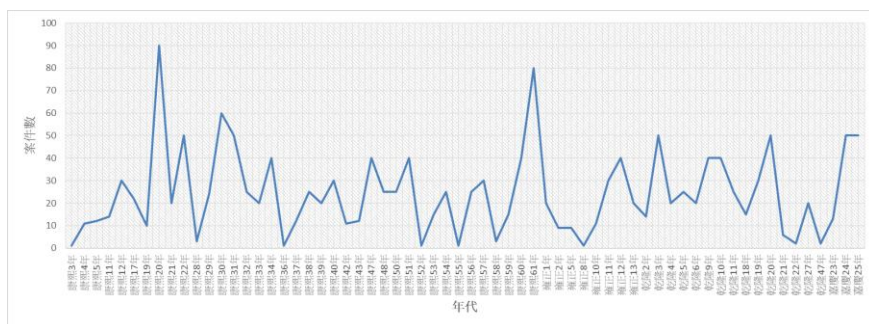


圖 2 各年度案件總數分布

四、各地案件數量分布

自地理分布圖中,可觀察到黃冊檔案案件有明顯的區域差異性,案件數量呈現由東部沿海向西部內陸遞減的空間特徵(表 3、圖 3)。圖 3 中以不同顏色區分案件數量級距:深紅色代表 150 件以上、橘色代表 100-150 件、藍色代表 15-100 件、綠色代表 15 件以下。標記為深紅色的地區中,以山東的 262 件高居首位,與江南(187 件)、廣東(165 件)並列為案件高頻區,除可能與人口密度及經濟活動頻繁有關外,亦需考慮地方司法申報密度與文書保存完整度之差異。作為賦稅與漕運重地,山東及江南行政體系相對完備,重案呈報與秋審制度之運作或更為規範,從而提高了相關

案件被納入黃冊的機率。換言之，數量分布亦可能反映司法行政網絡的密度，而非單純犯罪發生率的高低。次高頻地區（橘色）則集中於傳統農業核心區，包括山西（145 件）、湖廣（145 件）、河南（129 件）及浙江（127 件），或可顯示華北平原與長江中游流域的社會矛盾較為突出。陝西（126 件）與江西（112 件）等地點（藍色）雖屬中段區，但仍高於福建（78 件）、四川（54 件）等邊緣省份，而西南地區（雲南、貴州）及新開發的奉天（綠色）甚至低於 10 件。此犯罪案件數分布不僅與清代人口密度、經濟活動強度呈正相關，更可能折射出地方司法體系效能差異與移民社會的穩定性問題。值得注意的是，直隸作為京畿重地僅有 28 件案件紀錄。



圖 3 各地案件數量分布

表 3

各地案件數量統計表

地名	數量	地名	數量
山東	262	四川	54
江南	187	廣西	34
廣東	165	直隸	28
湖廣	145	貴州	26
山西	145	雲南	16
河南	129	安徽	6
浙江	127	奉天	3
陝西	126	江蘇	2
江西	112	湖南	2
福建	78		

五、各年度各籍貫與各旗案件分布

根據本研究所採黃冊檔案數據統計（圖 4），清代康熙至嘉慶年間各地案件數量分布呈現明顯的時空差異特徵。在空間分布上，直隸、山東、江南、浙江、江西、河南等核心行政區的案件記錄頻次與數量顯著突出，如康熙 20 年（1681）直隸錄得 31 例、山東 19 例、山西 24 例等，且全年度逾百例；康熙 30 年（1691）江南地區更出現 48 例的峰值。這些區域作為清代經濟與行政重心，其案件記錄的密集度反映了官方文書體系對核心地帶的監控強度。時間維度上，康熙朝自 20 年後進入案件記錄的高發期，特別是在康熙 20 至 31 年間（1681-1692），華北與江南地區持續出現大規模案件上報，其中直隸、浙江、山西等省多次達到雙位數記錄。雍正朝雖整體記錄稀疏，但雍正 12 年（1734）仍出現如安徽 24 例的集中記載。乾隆時期案件分布呈現擴散態勢，其中較高頻者，如乾隆 3 年（1738）安徽 38 例、江西 12 例，9 年（1744）廣東 8 例、河南 21 例、福建 13 例，1 年（1746）湖南 10 例、直隸 19 例，顯示案件記錄範圍向華南、華中延伸。嘉慶晚期廣東地區出現 56 例的階段性高峰，直隸與陝西分別維持 7 例與 12 例的記錄。這種空間分布格局或許也體現了清代檔案編製過程中的地域

覆蓋特徵，記錄波動既反映了檔案彙編的階段性重點，也可略為觀察不同時期司法文書制度的運作模式。

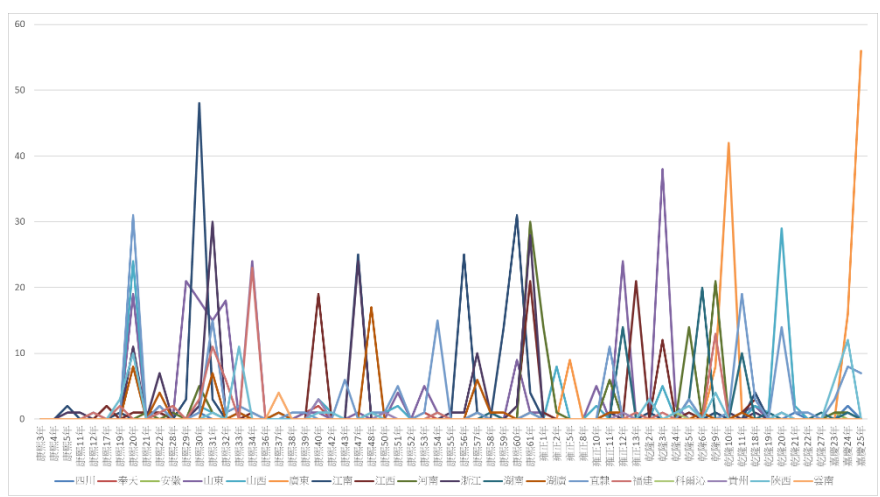


圖 4 各年度各籍貫案件分布

而將籍貫屬八旗之旗人分開觀察（圖 5），則可見康熙朝是旗人案件紀錄最集中的時期，雖因與漢人分屬不同審理機制，相對之下案件數極少，但仍可見部分案例，且於康熙朝前、中、後期各有高峰，這可能反映了當時朝廷對八旗管理的強化，而案件數量中又以正黃旗和鑲黃旗較為突出。進入雍正和乾隆時期後，案件數量顯著下降。乾隆中後期至嘉慶年間，許多年份甚至完全沒有案件紀錄，出現較長的斷層期。

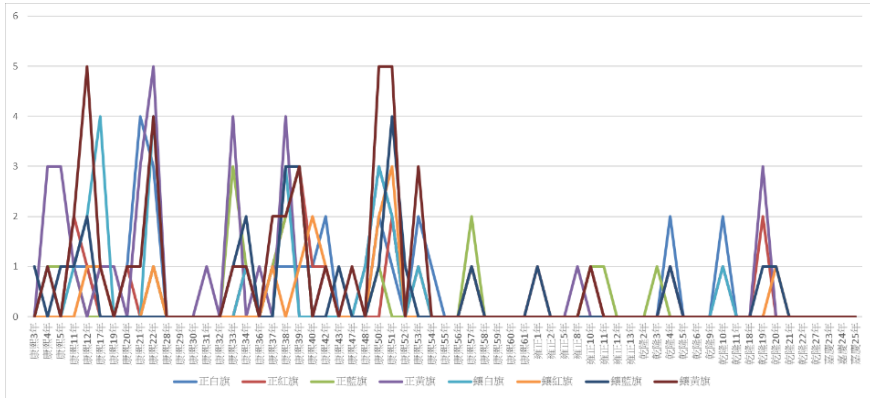


圖 5 各年度各旗案件分布

六、綜合討論

本研究結合文字探勘與詞頻分析，揭示清代黃冊中刑案紀錄在案件類型、犯罪動機、時空分布與族群差異上的若干結構性特徵。如犯罪類型分布顯示，黃冊中暴力犯罪（如殺人、毆傷）占據壓倒性比例，「打死」、「毆傷」為高頻詞彙；反之，性犯罪與財產犯罪記錄較少。此一分布與《吏部處分則例》中對命案、重盜案之查辦情況、審理時長等，均與政績考核密切相關（鄧建鵬，2019），導致基層官員傾向報送暴力案件以彰顯績效。性犯罪方面，「通姦」紀錄明顯高於其他性暴力事件，反映司法對違反儒家倫理秩序之特別關注，而非對性侵行為本身的譴責。

犯罪動機分析之標註與統計結果顯示，「家庭糾紛」與「感情糾紛」為最常見動機，其案件數超過「財務糾紛」，反映家庭內部為暴力事件好發場域。此現象未必直接反映家庭衝突在社會中最為頻繁，而可能與刑部呈報制度及司法論述慣例有關。誠然，人口快速成長所引發的家庭資源緊張，可能導致分家與贍養爭議激化；但對於涉及親屬間殺傷之案件，官員在呈報時往往需交代人倫關係，以說明犯案情節之輕重與情理判斷，或出於減輕刑責或利於和解等考量，使家庭關係成為敘事焦點。因此，此一分布結果或可理解為司法文本中「家庭倫理」作為責任歸屬與量刑判準的重要框架，而不僅僅是社會家庭衝突頻率的直接反映。

而案件時空間分布顯示，康熙二十年（1681）前後案件數顯著增加，除三藩之亂後社會秩序重整外，亦可能與中央行政節奏調整（如清理動盪期間積累之案件）有關。雍正朝下降至乾隆中期後再升，可能與人口壓力與訴訟文化擴張有關。空間上，山東與江南案件密集，可能反映其為交通與經濟中心、社會互動頻繁所致，但亦可能因文書保存完整與申報文化發達而產生偏差。西南邊陲地區因改土歸流未竟，司法相關記錄相對較少。

黃冊記錄所反映的犯罪樣態深受當代司法申報制度與治理邏輯所左右，其中高頻詞彙與統計分布，都需結合司法制度、社會風氣、地方背景等史實加以審視，本研究由特定主題檔案進行鳥瞰式整體分析，呈現檔案內容核心概念的分布，其意在於提供進一步史料研究的啟發，並非作為論斷史事的依據，也無意將分析結果視為社會實況的直接反映，研究價值在於探勘整體檔案關鍵資訊並揭示內容梗概，作為進一步挖掘探究特定研究議題的起點。

伍、結論

本研究以清代黃冊刑案紀錄為核心，結合 GAI 模型與數位人文分析技術，從語義分類、時空視覺化到方法論實踐，嘗試將非結構化的黃冊刑案紀錄轉化為可量化的研究資料，對黃冊資料進行系統性探索，並提出初步實證，觀察官方檔案敘事中呈現的犯罪類型與動機分布特徵。

一、語義分類與犯罪動機辨識

本研究運用 GAI 模型對黃冊文本進行語義分析，初步建立包括殺人、傷害、盜竊、性犯罪等主要類別的分類架構。模型能在多數情境中有效捕捉敘述中隱含的行為特徵與動機，展現一定的語義理解能力。然而，面對古典漢語中語義模糊或語境依賴較高的表述，模型辨識仍有限，顯示其準確性尚須透過辭彙標準化與人工校對加以補強。整體而言，GAI 模型具備輔助結構化歷史文本的潛力，可作為犯罪研究與歷史語料分析的初階工具之一。

二、犯罪活動的時空分布分析

結合時間資訊與地理標註，透過視覺化呈現康熙至嘉慶各朝的案件分布變化，描繪出清代前中期刑案的時空趨勢。康熙中期案件數量逐漸上升，可能與社會重建與地方秩序重整有關；雍正時期相對下降，或反映其整肅訟獄與推行吏治的成效；乾隆與嘉慶時期則再度增加，顯示隨人口壓力與社會流動性的提升，犯罪率亦呈現波動，但此一推論也僅限於本研究所分析之黃冊檔案量。地理分布方面，山東、江南地區為主要的案件集中區，西南與邊疆地帶案件數偏低，可能受限於司法制度覆蓋程度、檔案保存完整性與地方社會結構的差異，然此一時空視角分析目的在於提供犯罪活動與社會制度互動之關聯。

三、數位方法與歷史知識生產

本研究結合詞彙斷詞、語義分類與地理視覺化等技術，將非結構化的歷史紀錄轉化為可運算與可詮釋的資料資源，不僅提升文本資料的處理效率，也為歷史研究提供可參考的資料操作模式與觀察尺度。惟仍須注意，漢文黃冊內容與數位處理成果並非絕對客觀真實的再現，而是經制度偏好、保存選擇與技術條件層層形塑。數位工具應被視為歷史研究的輔助途徑，而非替代歷史研究的詮釋。鑒於黃冊檔案的樣本高度不均，且原始來源可能帶有蒐集目的的選擇性，本研所得的時空分布與類型統計結果，僅能視為數位方法應用於現存資料的一次探索性示範，非反映清代社會治安實況的實證性結論。

陸、未來研究建議

本研究透過語義分類與時空視覺分析，揭示黃冊刑案中反映的社會行為模式與制度邏輯，並初步觀察 GAI 與數位人文技術於清代文本中的應用效果。未來可朝以下方向持續拓展與深化：

一、進一步建立資料架構

可考慮以人物、地點與罪名等核心資訊為基礎，嘗試建立跨案件的資

料關聯，作為後續視覺化、網絡分析或制度研究的基礎。

二、嘗試跨文本對照與多語料參照

將黃冊內容與其他相關歷史文本進行初步比對，可觀察司法語言與案例描述在不同語境中的表現與變化，為後續文本詮釋與語料整合奠定方向。

三、加入人工分類檢視以驗證模型效度

透過人工標註一部分資料樣本，與 GAI 分類結果交叉比對，不僅可評估模型在歷史語境中的準確性與偏誤，更可作為優化語義分類模型的重要依據，提升分析結果的可靠性與學術說服力。

四、與歷史學者合作深化研究內容

本研究是由圖書資訊學者研究數位工具剖析歷史檔案內容的整體鳥瞰式分析，未來需要與歷史學跨域合作針對史學問題意識、方法有效性以及研究結果解釋加強探索，對於歷史研究議題較能提供專深的剖析結果。

（接受日期：2026 年 3 月 30 日）

參考文獻 References (Romanized & Translated)

- 于志嘉（2012）。明代軍戶中的家人、義男。《中央研究院歷史語言研究所集刊》，83(3)，507-570。
<https://www1.ihp.sinica.edu.tw/Publications/Bulletin/11/Article/47> 【Yue, C. C. (2012). Adopted family members within Ming dynasty military households. *Bulletin of the Institute of History and Philology, Academia Sinica*, 83(3), 507-570. (in Chinese)】
- 于志嘉（2018）。明代軍戶家族的戶與役：以水澄劉氏為例。《中央研究院歷史語言研究所集刊》，89(3)，541-604。
<https://www1.ihp.sinica.edu.tw/Publications/Bulletin/1056/Article/2013> 【Yue, C. C. (2018). Household registration and service obligations of military-household families during the Ming: A case study of the Shuicheng Lius. *Bulletin of the Institute of History and Philology, Academia Sinica*, 89(3), 541-604. (in Chinese)】
- 中央研究院近代史研究所（n.d.）。清代奏摺檔案系統。
<https://mhdb.mh.sinica.edu.tw/document/about.php> 【Institute of Modern History, Academia Sinica (n.d.). *Qing Dynasty Memorial Archives*. (in Chinese)】
- 衣若蘭（2024）。明代賦役改革前的婦女與戶籍簿冊登記。《臺大歷史學報》，73，43-98。
[https://doi.org/10.6253/ntuhistory.202406_\(73\).0002](https://doi.org/10.6253/ntuhistory.202406_(73).0002) 【Yi, J. L. (2024). Women and the household registration system in Ming China. *Historical Inquiry*, 73, 43-98. (in Chinese)】
- 李新峰（2010）。論明初里甲的輪役方式。《明代研究》，14，17-43。
<https://doi.org/10.29971/JMS.201006.0002> 【Li, X. F. (2010). The pattern of circulating corvee in the early Ming Li-Jia system. *Journal of Ming Studies*, 14, 17-43. (in Chinese)】
- 邱偉雲（2020）。中國近代天道概念的形成與演變—基於數位人文視野。《中國文哲研究通訊》，30(2)，31-68。
<https://www.litphil.sinica.edu.tw/newsletter/118/118-31-68.pdf> 【Chiu, W. Y. (2020). The formation and development of the concept of "Providence" in modern China from the perspective of digital humanities. *Newsletter of the Institute of Chinese Literature and Philosophy*, 30(2), 31-68. (in Chinese)】
- 胡其瑞（2024）。數位人文視角下《淡新檔案》中的教堂、教士與教民。《臺灣宗教

研究, 23(1), 159-187。

<https://drive.google.com/file/d/1pBm3IgLGPR077PZwN-cJcDO8AICqzOsJ/view>

【Hu, C. J. (2024). The churches, missionaries and believers in Danxin archives from the perspective of digital humanities. *Taiwan Journal of Religious Studies*, 23(1), 159-187. (in Chinese)】

國立北平故宮博物院文獻館 (1936)。內閣大庫現存清代漢文黃冊目錄。台聯國風出版社。【Department of Manuscripts, Palace Museum of Peiping (1936). *nei ge da ku xian cun qing dai han wen huang ce mu lu*. tai lian guo feng chu ban she. (in Chinese)】

國立北平故宮博物院文獻館、國立北京大學文科研究所、國立中央研究院歷史語言研究所 (編) (1947)。清內閣舊藏漢文黃冊聯合目錄。國立北平故宮博物院。【Department of Manuscripts, Palace Museum of Peiping, Institute of Humanities of Peking University, & Institute of History and Philology, Academia Sinica. (Ed.). (1947). *qing nei ge jiu cang han wen huang ce lian he mu lu*. Palace Museum of Peiping. (in Chinese)】

許雅惠 (2018)。北宋晚期金石收藏的社會網絡分析。新史學, 29(4), 71-124。

<https://www.newhistory.org.tw/contents/2022032511360648082/16502832825117jsessionid=0B51A2719221777AE72AA0A523083D7E> 【Hsu, Y. H. (2018).

The social networks of antiquities collectors in the late northern song. *New History*, 29(4), 71-124. (in Chinese)】

陳冠至、陳柏溢、黃敬程 (2018)。孰執牛耳？明代蘇州藏書家社群的數位人文解析。教育資料與圖書館學, 55(3), 243-284+i-x。

[https://doi.org/10.6120/JoEMLS.201811_55\(3\).0019.RS.AM](https://doi.org/10.6120/JoEMLS.201811_55(3).0019.RS.AM) 【Chen, K. C., Chen, B. Y., & Huang, C. C. (2018). Who was the key figure? A social network analysis of suzhou book collector groups in Ming dynasty through a digital humanities approach. *Journal of Educational Media & Library Sciences*, 55(3), 243-284+i-x. (in Chinese)】

陳柏溢、陳冠至 (2021)。清代紹興府藏書家之社會網絡分析。圖資與檔案學刊, 98, 83-131。https://doi.org/10.6575/JILA.202106_(98).0003 【Chen, B. Y., & Chen, K. C. (2021). A social network analysis of shaoxing book collector groups in Qing dynasty. *Journal of InfoLib and Archives*, 98, 83-131. (in Chinese)】

趙思淵、潘芸淇 (2024)。從史料量化到數位人文：明清社會經濟史研究中的數位

- 方法演進。《漢學研究通訊》，43(1)，13-21。【Zhao, S. Y., & Pan, Y. Q. (2024). From quantification to digital humanities: The evolution of research approaches to the social-economic history of the Qing Dynasty. *Newsletter for Research in Chinese Studies*, 43(1), 13-21. (in Chinese)】
- 趙郁楠（2018）。清代內閣滿文黃冊的整理與利用。《歷史檔案》，2018(3)，140-144。https://fhac.com.cn/research_institute/history_detail/detail/4641.html
- 【Zhao, Y. N. (2018). Qing dai nei ge man wen huang ce de zheng li yu li yong. *Historical Archives*, 2018(3), 140-144. (in Chinese)】
- 劉錚雲（1998）。舊檔案、新材料——中研院史語所藏內閣大庫檔案現況。《新史學》，9(3)，135-162。
- https://www.newhistory.org.tw/contents/202203251120000495/1649601222349940;jsessionid=D048154F006FAD5C7156DCE661CA7F19 【Liu, C. Y. (1998). The grand secretariat archives at the institute of history and philology, academia sinica. *New History*, 9(3), 135-162. (in Chinese)】
- 鄧建鵬（2019）。清代州縣詞訟積案與上級的監督。《法學研究》，41(5)，173-190。
- https://faxueyanjiu.ajcass.com/Magazine/Show?ID=94031 【Deng, J. P. (2019). Civil case backlog at the county level and supervision over it by superior officials in the Qing dynasty. *Chinese Journal of Law*, 41(5), 173-190. (in Chinese)】
- 關伯丞、宋茆熏、沈錕坤、蔡銘峰（2012）。由史料中探勘職官年表：以清聖祖實錄為例。載於國立臺灣大學（主編），《第四屆數位典藏與數位人文國際研討會論文集》。國立臺灣大學出版中心。【Que, B. C., Song, F. S., Shan, M. K., & Tsai, M. F. (2012). Mining official chronology from historical documents: By example of veritable records of the Qing Kangsi. In National Taiwan University (Ed.), *4th International Conference of Digital Archives and Digital Humanities*. National Taiwan University Press. (in Chinese)】
- Digital Panopticon. (2020). *Tracing London convicts in Britain & Australia, 1780-1925*. https://www.digitalpanopticon.org/
- Guiney, T. (2020). Excavating the archive: Reflections on a historical criminology of government, penal policy and criminal justice change. *Criminology & Criminal Justice*, 20(1), 76-92. https://doi.org/10.1177/1748895818810333
- Hsiang, J., Chen, S. P., Ho, H. I., & Tu, H. C. (2012). Discovering relationships from imperial court documents of Qing China. *International Journal of Humanities*

- and Arts Computing*, 6(1-2), 22-41. <https://doi.org/10.3366/ijhac.2012.0036>
- Mao, L., & Ma, Z. (2012). Writing history in the digital age: The New Qing History Project and the digitization of Qing archives. *History Compass*, 10(5), 367-374. <https://doi.org/10.1111/j.1478-0542.2012.00841.x>
- The Digital Humanities Institute of University of Sheffield. (2023). *The Proceedings of the Old Bailey: London's Central Criminal Court, 1674-1913*. <https://www.oldbaileyonline.org/>
- Park, N. (2020). Officials and Chinese justice: Public and private wrongdoing in Qing law. *T'oung Pao*, 106(5-6), 661-713. <https://www.jstor.org/stable/27067795>