

人工智慧合同審查的知識圖譜構建

張 泳¹

【提要】弱人工智慧時代，人工智慧與司法融合日益加深，而司法人工智慧的運作與應用有賴於知識圖譜的構建。現有知識圖譜存在顆粒化程度較低、實體要素不夠多元等問題，且多集中于案件裁判、法律問答等方面。筆者結合當前人工智慧發展形態和實務所需，認為應當從民商事領域入手，構建有關合同整體性審查的知識圖譜。在梳理法律法規、司法解釋等法律規範的基礎上，結合指導性案例和法學通說，拆分、提煉要素實體並搭建其間的關係，完成“知識整理——要素加工——關係連結——資料存儲”的知識圖譜構建過程。通過在龐大的法律文本資料中厘清實體、實體屬性和實體之間的關係，構建知識圖譜，搭建智慧合約模型，並以網路交易合同為例進行運用，完成合同審查的人工智慧司法應用。

【關鍵字】弱人工智慧；知識圖譜；合同審查；智能司法

一、問題的提出

Web 3.0 時代，司法人工智慧成為法律研究和實務的風向之一。現有的人工智慧技術仍處於“弱人工智慧”階段，存在“可解釋性”的瓶頸，實質上僅是提取不同案件的共同要素構建知識圖譜，將既有規則或知識圖譜以電腦程式轉化為自動執行或推送的操作流程，² 因此知識圖譜是司法人工智慧進行演算法建模、識別案件的基礎，是司法人工智慧得以運作的前提，關乎人工智慧的應用效果。

1 *張泳，橫琴粵澳深度合作區研究院特聘研究員，中央司法警官學院司法人權研究中心研究員，澳門大學法學博士。

2 顧全：《數位經濟案件分類體系及裁判規則研究》，載《中國應用法學》2022年第5期。

2022 年 12 月 9 日，最高人民法院發佈《最高人民法院關於規範和加強人工智慧司法應用的意見》（以下簡稱《意見》）。³《意見》堅持倫理道德原則的人工智慧司法應用理念，指出總體目標是到 2025 年基本建成較為完備的司法人工智慧技術應用體系，到 2030 年建成具有規則引領和應用示範效應的司法人工智慧技術應用和理論體系，提出人工智慧司法應用需遵循五項基本原則。即應用司法人工智慧技術應當遵循安全合法原則、公平公正原則、輔助審判原則、透明可信原則、公序良俗原則，這將進一步推動人工智慧與司法裁判工作相融合，全面深化智慧法院建設，推動數位正義的高水準發展，更是推動智慧法治建設走上一個新的臺階。

知識圖譜作為司法人工智慧的主流演算法，通過對自然語言的識別、理解，分析其中所含語義並處理歧義，搜索、組織及展示知識，促進開放協作、溝通聯結，因此被越來越多科研單位和科技公司採用。但是，我國缺乏官方的、權威的司法知識圖譜。完整的司法知識圖譜至少應該包括個案知識圖譜、證據知識圖譜和案由邏輯知識圖譜三個方面。⁴現有的許多知識圖譜僅是針對於簡單的案件構成進行視覺化的圖像描繪，資料、模型的顆粒化程度較低，考慮的要素與實體不夠詳細全面，使得司法人工智慧預測裁判的準確率較低，無法發揮真正的“智慧”作用。法律知識圖譜應該如何構建與應用？法律知識圖譜如何助力於弱人工智慧的研究與發展？本文擬對此問題作出探索與嘗試。

二、法律知識圖譜概述

為研究知識圖譜的構建方式以及在司法領域的具體應用，需要先梳理知識圖譜的概念、特徵及現狀，探究在法學領域引入知識圖譜的必要性與可行性，分析司法領域現有知識圖譜開發與應用概況，從中獲得有關知識圖譜構建的思路。

（一）知識圖譜相關概念與理論

知識圖譜（Knowledge Graph）是基於對語義網（Semantic Network）標準與技術進行研究的產物，語義網的概念在 20 世紀中葉被提出，通過主張推廣、完善使用本體模型，以形式化地表達資料中隱含的語義，⁵在此基礎上產生了諸如 RDF 模式和萬維網本體語言等模型。2012 年，穀歌首創性地提出了知識圖譜的概念，正式開啟了知識圖譜的應用。就分類而言，知識圖譜屬於語義網的一種；論其本質，知識圖譜相當於以圖資料為主要構造的知識庫；究

3 《最高人民法院關於規範和加強人工智慧司法應用的意見》，法發〔2022〕33 號，2022 年 12 月 9 日發佈。

4 王竹：《司法人工智慧推理輔助的“准三段論”實現路徑》，載《政法論壇》2022 年第 5 期。

5 See BERNERS-LEE T. ,HENDLER J. ,LASSILA O. ,The semantic Web, Scientific American Magazine, 1-4 (2008) .

其內容，知識圖譜是一種用於表達實體間關係的語義網路；觀其形態，知識圖譜以形式化的節點（Point）和標注的邊（Edge）來呈現。穀歌以此提高搜尋引擎的效率，給用戶帶來更智慧的搜索效果和體驗，自此逐漸在學術界和實務界普及，被廣泛運用於智慧問答系統、個性化推薦、語義搜索、知識推理、情報資訊分析等多種領域，形成當前拓寬應用服務的研究熱點，知識圖譜的概念也有所擴充，逐漸成為各種類型大規模知識庫的泛稱。

就其構成而言，知識圖譜包括資料層（data layer）和模式層（schema layer）兩部分，前者是知識圖譜的核心部分；後者是知識圖譜的概念模型和邏輯基礎，通過本體定義的規則和公理，以此規範和約束資料層。知識圖譜可以分為早期由專業人員手工構建的小規模圖譜，以及依賴於互聯網資訊自動提取技術提煉出的實體及其關係而構建的大規模圖譜；可以分成日常的、廣泛的通用知識圖譜與特定的、準確的行業知識圖譜。⁶ 本文所討論的合同審查知識圖譜，屬於法律領域專業化的行業知識圖譜。基於此，有學者總結了知識圖譜的特點：（1）主要描述真實世界中的實體以及實體之間的語義關係，並以網路結構進行組織和表示（2）使用一種模式（Schema），來定義實體的類別和關係（3）允許任意實體之間的相互關聯，強調此種關聯及屬性值（4）覆蓋各種主題領域，側重來源差異化的知識之融合與過濾。⁷

（二）知識圖譜引入司法的必要性與可行性

司法人工智慧的運作效率很大程度上取決於資料的信度和效度。資料的信度是針對其真實性和可靠性的評估。作為動態化的集合，司法知識圖譜將法律法規、司法檔、法院判例、證據規則和案件事實融合於一體，⁸ 基於司法公開化程度較低的當下現狀，許多資訊存在較高的獲取門檻，即使是公開的司法資訊，出於各種因素的考慮或是受限於資料本身表達與解讀的有限性，其中也存在非真實的部分，存在著“表像性”特徵，外界獲取的司法資訊是用以證明法律決策正確性而按照一定程式製造出來的，難以充分、真實地反映法律決策中所真正被考量的實質因素，⁹ 因此需要人類通過知識圖譜以貼標籤的形式進行預處理，以一般的經驗排除明顯偏離常態的失真資料，挖掘“表像性”資料背後的真實資訊，提煉“實質性”資料中存在的普遍標準。資料的效度則反映了資料的準確性，是司法人工智慧能夠得出正確結論的基礎。然而由於中文語義的豐富性、表述的多樣化和理解的主觀性，司法資料的結構化程

6 黃恒琪、于娟、廖曉、席運江：《知識圖譜研究綜述》，載《電腦系統應用》2019年第6期。

7 徐菁：《面向中文知識圖譜的開放式文本資訊抽取關鍵技術研究》，國防科技大學2018博士學位論文。

8 馬長山：《司法人工智慧的重塑效應及其限度》，載《法學研究》2020年第4期。

9 左衛民：《關於法律人工智慧在中國運用前景的若干思考》，載《清華法學》2018年第2期。

度較低，在所指物件同一但表述方式不同的情況下，法律人工智慧無法精準識別並歸為同類；再者，由於實務中不同機關單位對於司法資料處理的標準存在差異，很可能人為地把同一類型資料歸入不同的處理機制，要求各地遵照同等標準進行司法資料處理顯然過於理想化，在當前司法地方性差異難以消解的情況下，司法人工智慧的有效運作很大程度上需要專業人士進行司法資料的篩選、清洗、歸類，將法律解構成數個關聯的邏輯模型，依賴於構建知識圖譜的“標籤”，以方便人工智慧的識別和深度學習。即使司法資料的信度和效度達到了要求，但是此部分資料在全部案件中的比重是否能夠作為代表性樣本反映司法裁判的基本規律，還有賴於人工的把控。更進一步地，在此類資料之外，還有大量的司法活動未被加以資料化的形式表達，如果沒有知識圖譜對部分資訊進行結構化、層次化梳理，司法的資料性特徵和非資料化部分雜糅在一起，人工智慧更加難以發揮用武之地。

就案件判決書與知識圖譜的關係而言，知識圖譜可以通過梳理案件中的重要實體節點及其關係，如涉案主體關係、簡單的法律因果關係、涉案物品與地點等，加上不同開發方向各自側重的類型化實體，擴充知識圖譜中的“邊”，從而將孤立的個案連結成為關聯查詢、檢索分析的智慧化網路，並進行直觀化的展示和描繪。在科技工具和犯罪手段多元化的今天，各式各樣的新興案件不斷出現，如果沒有知識圖譜的結構化顯示，未來出現的新實體將不易被歸類；另設類別則浪費了原先的工作基礎，需要耗費更多的成本；機械添加則有損模型的顆粒度，影響人工智慧運作的效果。而應用知識圖譜的優勢在於，圖譜基本的節點和邊的框架一旦構建，其外延具有規模化的擴充潛力，為新實體及其關係的納入和拓展留有充分的空間，便於從一般到具體的要素提煉、環節拆分和資料分層。

知識圖譜的構建是司法人工智慧裁判系統進行學習的基礎，以此獲得自主處理新知識的能力，這是類案裁判的邏輯。因此，整理知識圖譜作為司法人工智慧系統研發的重要環節，相當於建立了司法人工智慧與外界錯綜複雜的資訊之間的橋樑，且橋樑上具有分岔口，可以過濾不同類型的資料資訊，給予人工智慧“精細化”的投餵養料。而司法人工智慧依靠知識圖譜中分門別類的“養料”，伴隨對於大量案件的學習過程，逐漸形成“案件要素——圖譜實體”、“案件關係——圖譜關係”的匹配能力，從而能夠理解、推理案件並進行自主裁判。

（三）現有法律知識圖譜的研究與應用

知識圖譜在法律人工智慧中起著基礎性的作用，有學者指出，理想情況下，法律人工智慧將類推學派的知識圖譜和聯結學派的深度學習相結合，實現機器學習。¹⁰具體到實務案件中，案件知識圖譜就是用符號形式描述的法律案件要素相互關係的——在類案推薦與偏離預警的

10 左衛民：《關於法律人工智慧在中國運用前景的若干思考》，載《清華法學》2018年第2期。

應用場景下，主要是案件情節與裁判結果之間關係的——結構。¹¹ 當前，人工智慧司法通過構建知識圖譜和演算法模型，在法律法規檢索、法律知識與類案推送、裁判結果預測、文書自動生成、法律智慧問答、司法資料視覺化等方面都有所涉及，實現了線上一鍵化開啟包括開庭到判決執行的全部流程，即“運用大資料與訴訟建模相結合的方法，對管轄法院、訴訟請求、賠償數額計算、法律依據援引等訴訟事項進行綜合化、模組化處理，由當事人選擇完成起訴和應訴程式”。¹²

國內開展了關於知識圖譜的諸多實踐，江蘇法院創造性地將刑事類案證據的知識圖譜嵌入“同案不同判預警平臺”，“成功預警高偏離度案件（三級預警）120 餘例，預警案例總數占案例總數 3.3%，準確率高達 92%”。上海“206 系統”工程採用了全國法院“眾籌”模式，通過人工和自動相結合的標注方式，來完成證據知識圖譜的繪製。¹³ 北京市“審法官”系統以法律構成要件為基礎，結合專業知識和案例經驗，依託基礎設施層獲取海量的司法資料，構建了完整的知識圖譜。法狗狗將知識圖譜與法律 NLP 相結合，為智慧諮詢提供解決方案。我國其他法院和檢察院，如蘇州法院的“智慧審判蘇州模式”、河北法院的“智審”等，都將法律知識圖譜應用於實踐，結合語義分析、語言處理等方法，人工智慧可自動抽取比對公訴意見書、起訴書、答辯狀中辯訴意見及事實理由，以概括案件情節和爭議焦點。¹⁴ 國外部分公司在法律風險評估和裁判預測領域運用了知識圖譜技術，以 Premonition 公司和 Hodge Jones & Allen 公司的審判結果預測系統為代表。¹⁵ 在學界，國內外法律知識圖譜的相關研究覆蓋了民事、刑事、行政等多個領域，包含交通事故、網路詐騙、經濟犯罪、國際刑事司法、人口販賣、刑事兇殺等不同類型的案件。秦永彬團隊以貴州省高級人民法院司法資料為基礎，提出一種以知識圖譜為核心的司法資料融合技術，構建多細微性實體和多層次知識圖譜，進一步進行知識進化學習與推測實體間關係。¹⁶ 洪文興等學者設計了針對交通事故案件的案情知識圖譜自動構建流程，以數十萬份交通事故案件判決書為樣本，研究為人工智慧在司法中的運用提供

11 王祿生：《司法大資料與人工智慧開發的技術障礙》，載《中國法律評論》2018 年第 2 期。

12 鄧恒：《人工智慧技術運用與司法創新》，載《人民法院報》2017 年 12 月 15 日，第 2 版。

13 馬長山：《司法人工智慧的重塑效應及其限度》，載《法學研究》2020 年第 4 期。

14 葉鋒：《人工智慧在法官裁判領域的運行機理、實踐障礙和前景展望》，載上海市法學會主編：《上海法學研究》第 5 卷，[出版者不詳]2019 年版，第 373-392 頁。

15 朱福勇、劉雅迪、高帆、王凱：《基於圖譜融合的人工智慧司法資料庫構建研究》，載《揚州大學學報》2019 年 11 月 29 日，第 6 版。

16 秦永彬、馮麗、陳豔平、黃瑞章、劉于雷、丁紅發：《“智慧法院”資料融合分析與集成應用》，載《大資料》2019 年第 3 期。

語義支援的有效途徑。¹⁷ 黃茜茜、楊建林以網路詐騙案件的司法判決書為研究物件，構建網路詐騙領域知識圖譜，並以此為基礎提出關於智慧司法知識服務框架的設計理念，拓展案件知識圖譜運用的實體類型與場景。楊陽選取某地下錢莊案的資料作為研究樣本，基於知識圖譜設計出涉眾型經濟犯罪知識建模方法，並對相關目標物件的社交網路與社會關係管道進行圖譜化的分析。¹⁸ 王雲才借助 Cite Space II 軟體，對犯罪學研究現狀、前沿和未來進行視覺化分析，¹⁹ 對國際刑事司法研究注重理論和實證的趨勢進行描繪。²⁰ 許振亮、劉喜美繪製了電信詐騙領域的演進知識圖譜和聚類知識圖譜，闡明了電信詐騙研究的演化路徑，分析了 7 個代表不同知識群體的聚類。²¹ 利用案件知識圖譜，Elezaj O 設計出一個智慧查案分析框架，該框架具有作案特徵挖掘、案件自動分類和犯罪網路視覺化等功能。²² Szekely 等構建了有關人口販賣的知識圖譜，結合相關司法資料進行研究，為打擊販賣人口執法活動提供幫助。²³ Pandey R 基於兇殺案件調查年表構建知識圖譜，以呈現證人、嫌疑人、偵探和證據四類實體以及它們之間的關係。²⁴

儘管法律知識圖譜的應用和研究在國內外均有所發展，但當前的知識圖譜多以案件、涉案人員、法院審判人員為實體節點構建其間的關係，較少涉及案件審判中的法律法規、量刑幅度、判決結果等實體，圖譜應用中的實體類型和場景不夠豐富，資料模型不夠精細，儘管類別多樣，但智慧裁判的效果較為一般。此外，現有研究多僅考慮其在知識推理和表達方面的應用，忽視對於法律文本表示和補充的作用。從智慧學習的角度而言，現有圖譜的研究基礎多來源於歷史資料庫，在搭建司法人工智慧知識圖譜的過程中，缺乏跟進新興司法動態、更新司法資料庫的主動學習經驗。

17 洪文興、胡志強、翁洋、張恒、王竹、郭志新：《面向司法案件的案情知識圖譜自動構建》，載《中文資訊學報》2020 年 1 月 15 日，第 1 版。

18 楊陽：《基於知識圖譜的涉眾型經濟犯罪知識建模與情報分析》，載《中國人民公安大學學報》2020 年 5 月 15 日，第 2 版。

19 王雲才：《犯罪學研究知識圖譜分析》，載《中國刑事法雜誌》2011 年第 10 期。

20 王雲才、張民：《基於知識圖譜的國際刑事司法研究視覺化分析》，載《上海公安高等專科學校學報》2014 年 6 月 20 日，第 3 版。

21 許振亮、劉喜美：《電信詐騙研究的知識圖譜分析》，載《中國刑警學院學報》2017 年 6 月 20 日，第 3 版。

22 See Elezaj O. , Yayilgan S. Y. , Kalemi E. , et al , Towards Designing a Knowledge Graph-Based Framework for Investigating and Preventing Crime on Online Social Networks , in Switzerland:Springer ed. , International Conference on e-Democracy.Cham , 181-195 (2019) .

23 See SZEKELY P. , KNOBLOCK C. , JASON S. , et al , Building and using a knowledge graph to combat human trafficking , The Semantic Web-ISWC , 205-221 (1964) .

24 See Pandey R. , Text Mining for Social Harm and Criminal Justice Applications , West Lafayette , United States:Purdue University , 2020.

三、合同審查與知識圖譜的內在契合

結合上述知識圖譜應用於司法人工智慧的現狀、當前弱人工智慧的發展形態以及審判實務中法院積壓案件的主要類型，基於下列考慮，筆者提出關於民商事合同整體流程審查的知識圖譜，以此推進締約方之間起草和談判合同的進程，輔助判斷合同的效力。

（一）針對民商事案件構建知識圖譜具有可行性

由於刑事案件的構成犯罪的標準、證明標準、審查標準等較高，且犯罪手段方式多元化不易歸類，需要對不同的證據作出更加細緻的鑒別，因此難以為知識圖譜的資料處理提供規範、統一的預設標準；而民商事案件中存在部分具有模式化、標準化的領域，如小額借貸、網路交易、交通事故、婚姻繼承、勞動糾紛等，其中的實體及其間的關係往往較為簡單明確，案情和證據較易把握，能夠在知識圖譜中以清晰明瞭的節點和邊的形式反映。另一方面，民事訴訟和刑事訴訟的模式存在不同，前者是當事人雙方的平等對抗，對於證據的披露與展示較為積極，知識圖譜能夠從中獲得更多“實質性”的資料和資訊；後者中控辯雙方地位天然不平等，控方的證據具有絕對效力，使得相關資訊的披露存在真實性和充分性方面的缺陷，司法人工智慧根據這些“表像性”的資料和資訊所推斷出的結論效果欠佳。

（二）針對合同爭議部分構建知識圖譜具有優越性

相比案件判決書，合同作為資訊資料的複合載體具有如下特徵：首先，合同中存在較多結構化條款，以小額借貸合同為例，合同中的數額、價款、標的、期限等較為明確，足以為知識圖譜提供“可解釋性”的語言；而案件判決書的內容本身較為複雜，特別是裁判說理部分存在大量非結構化的表述，構建知識圖譜時需花費大量人力、物力用於分類篩選。此外，案件判決書文本的撰寫來自審判人員，不可避免地具有個人心證和主觀判斷；而這些簡單民事案件的合同起草具有通用範本可以參考，具有固定的要素和統一的表述，特別是當合同締約雙方均為非法律專業人士時，合同很大程度上依賴於範本本身，因此利用知識圖譜內含的專業知識來填充模式化的合同，無需用戶自己填寫所有細節，能夠提高合同締約的品質，提高民商事活動的效率。上述所言的模式化、標準化小額類型案件，其中的法律要素和關係基本上反映在合同之中，通過對合同部分的分析，能夠較高效地提取構建知識圖譜所需的實體及關係。

（三）針對合同整體流程之審查構建知識圖譜具有科學性

合同審查的主要物件為合同文本和條款，比起傳統的表格模式或者文字模式，知識圖譜能夠以更加靈活的互連方式存儲資料，這對合同而言是至關重要的——合同條款之間的排列組合與上下文邏輯關係，可能影響文本的語義分析和文義解釋；合同中的某些條款具有的推定含義，影響其他條款的關聯性解讀。通過知識圖譜將這些實體及其關係進行整合，將相關

資料資訊存儲於知識圖譜中，能夠簡化合同流程，並使使用者能夠更便利地獲得資料。進一步地，對合同整體流程的審查，包括合同締約主體、締約程式等，能夠使其中的各項要素相互印證，以增強審查環節的完整性。且合同流程整體涉及的資訊資料，也符合司法人工智慧要求的可解釋、易獲得、充分真實等特徵，具有構建知識圖譜的可行性。

四、合同審查知識圖譜的具體構建

構建一個知識圖譜需要四步，即知識體系構建、知識抽取、知識融合及知識存儲。²⁵ 相應地，結合司法人工智慧合同審查的特點，合同審查知識圖譜的構建過程亦可以被大致分為四個步驟：一是構建知識體系，也即知識整理，以獲得完整的合同合規知識體系；二是知識抽取，以抽取需求的目標資料資訊；三是要素加工與關係連結，以獲取某個實體或概念的完整表述。四是知識圖譜存儲，以高效存儲上述處理好的知識資料。另外，從合同審查知識圖譜的構建方法上來看，因法律領域不同於通用領域有極高的規範要求，故應考慮自上而下的構建方法，先構建知識體系，也就是合同審查知識圖譜的模式層。²⁶ 具體構建流程如圖 1 所示。

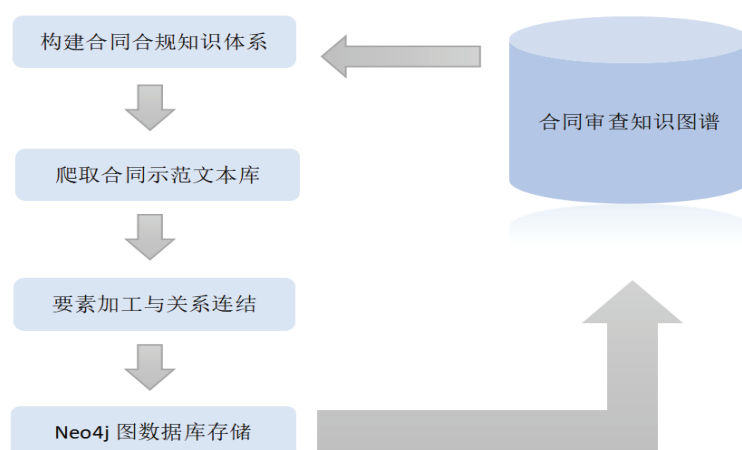


圖 1 合同審查知識圖譜構建架構

25 劉則淵、陳悅、侯海燕：《科學知識圖譜：方法與應用》，北京人民出版社 2008 年版。

26 知識圖譜主要有兩種構建方法：（1）自上向下的構建方法，即先構建知識體系即模式層，包括醫療實體的類別、醫療實體的屬性以及醫療實體之間的關係；然後通過將醫療資料與知識體系相映射來構建知識圖譜。醫療領域常採用這種構建方法，因為該領域資料需要很高的精度、品質及行業的各種規範要求；（2）自下向上的構建方法，先通過知識抽取技術從多源異構資料中提取實體、屬性和關係；然後將這些實例知識資料進行總結歸納，逐層向上抽象為概念，構建知識體系，形成模式層。例如百度知識圖譜，包含大量資料，強調資料的廣度。這種構建方式不依賴於人工，自動化程度高，可以構建出更為豐富全面的知識體系，但知識雜訊大，準確性不高，常用於通用領域的知識圖譜的構建。參見趙軍、劉康、何世柱：《知識圖譜》，北京高等教育出版社 2018 版。

（一）知識整理

構建知識圖譜的第一步是對於該領域相關知識的系統整理。

一是梳理適用於裁判的法律法規和司法解釋。之所以從法條規範出發，有如下考慮因素：此部分作為該領域的基礎性規範具有指導性的作用，是應對類案通行適用的標準，也是各地司法機關進一步解釋說理的出發點，在處理相關案件時能夠進行較為統一的處理，以此為知識圖譜的構建提供模式化的預設標準；比起案例，多數法律規範本身包含了許多要素，在立法時就有明確的指向，符合“可解釋性”的特徵，後續整理提取的難度較小，十分適合具有初步學習的弱人工智慧。

二是梳理該領域的案例，特別是最高人民法院發佈的具有指導性作用的裁判文書。現有立法中不可避免地存在規定模糊、存在歧義的部分，即使規則清晰、指示準確，也會隨著新情況的出現使得原有用詞的所指內涵有所變化，使得原先的文義解釋逐漸難以適用。相比於指導性案例，司法解釋存在一定的滯後性、抽象性、依附性，難以應對實務層出的各種問

題。因此對案例判決書進行梳理是構建知識圖譜的擴展性環節，有助於增加知識圖譜中實體的種類，對實體之間的多元化關係進行補充。再者，案例本身是“資料資訊整理——知識圖譜構建——機器自主學習”框架中必不可少的血肉，司法人工智慧根據對應知識圖譜中的要素進行案例拆解的過程，正是實現從知識輸入、吸收轉化到自主學習、輸出結果的關鍵步驟，決定了司法人工智慧的“智慧”程度。而在知識圖譜中梳理指導性案例判決，相當於把這一過程部分前置化，即把要素完整、規則清晰、最具有指導性和代表性的最高院指導性案例放入知識圖譜的構建過程，與司法解釋一起梳理，輔助其他法律淵源的應用；把大量的其他案例判決歸入後續的機器自主學習階段。即在知識圖譜初步構建的基礎上，以大量其他案件作為深度智慧學習的主要對象，以此增強司法裁判機器的智慧化程度。

三是梳理與上述內容契合的法學通說。有學者指出，個案知識圖譜基於實體法律規則構建，證據知識圖譜基於證據法律規則構建，案由邏輯知識圖譜基於法學理論通說構建。²⁷可見，法學理論通說也是構建知識圖譜的基礎之一。2018年，最高人民法院《關於加強和規範裁判文書釋法說理的指導意見》第13條規定，法官可以運用“法理及通行學術觀點”作為論據，論證裁判理由，提高裁判的可接受性。意見明確肯定了法學理論通說在案例論證中的作用，儘管相比於德國法傳統中的通說制度，我國法學通說的司法應用存在爭議。指導性案例的拘

27 高翔：《人工智慧民事司法應用的法律知識圖譜構建——以要件事實型民事裁判論為基礎》，載《法制與社會發展》2018年第6期。

束力來源於事實上的權威，法學通說對於司法裁判的拘束力則源自其知識上的權威。²⁸ 在上述制定法存在漏洞、空缺，或是表述語義不清、規則不明時，在指導性案例涵蓋情形有限不足以解決問題時，法學通說可以在前兩者的基礎上，代表法學知識和裁判經驗上的共識，凝結支持正當性判決的法理，因此梳理法學通說具有必要性。儘管梳理主流法學理論在構建知識圖譜的過程中存在較大難度，但相比於前兩者，並非構成知識圖譜的主要部分，僅是起補充作用，內容、占比上均較少，因此該部分所費成本仍保持在可控範圍內。

（二）知識抽取

在龐大的法律文本資料中抽取實體、實體屬性和實體之間的關係是知識抽取的主要內容。在合同審查方面，爬取國家市場監督管理總局開發的合同示範文本庫以充實知識圖譜的資料庫。²⁹ 具體而言包含以下三個任務：

一是實體識別與屬性分析。命名實體識別是自然語言處理（Natural Language Processing，NLP）的一個基礎任務，其可以識別出語料中的組織機構名稱、人名及地名等命名實體。另一方面，知識抽取就是要從非結構化的文本中抽取結構化資料與特定的關係。例如：“小明和小李於 2022 年在北京大學簽訂了一份買賣合同”，其中“小明”“小李”“2022 年”“北京大學”“買賣合同”都是命名實體，而這些實體包含了合同當事人、簽訂時間、有名合同類型等重要資訊，其他詞語將這些實體串聯起來才能有完整的意思表示。具體而言，合同審查知識圖譜中的實體包括合同類型、合同當事人及第三人。其中，合同類型亦是合同審查知識圖譜的核心概念，為了對合同進行充分描述，突出合同的重點資訊，在知識圖譜中的合同類型實體屬性包括相關法條、司法解釋及理論通說；合同當事人包含甲方與乙方，一般包含的屬性有公司名稱、姓名、出生日期、成立時間、行為能力、權利能力；第三人包括有獨三與無獨三，而其實體屬性同前述合同當事人的。

二是關係抽取。知識圖譜中的關係是指兩個或多個實體之間的某種聯繫，而關係抽取就是要實現自動地從語料文本中檢測並識別出實體之間的某種語義關係。通常，關係抽取的結果將被表示為一個三元組（實體 1，關係，實體 2）。在關係抽取方法的種類上，常用的有基於機器學習的關係抽取於基於規則的關係抽取。其中，基於機器學習的關係抽取方法，按照機器學習方法對語料庫的不同需求，可以被大致分成三類，包括有監督關係抽取、無監督關係抽取和弱監督關係抽取。其中，有監督關係抽取方法是指使用人工標注的語料進行訓練。

28 周曉帆：《論法學通說在我國司法實踐中的應用——以裁判文書說理為視角》，載《西部法學評論》2021 年第 6 期。

29 截止 2022 年 6 月 16 日，文本庫已收錄各類合同示範文本近 500 種。國家市場監督管理總局官方網站，<https://www.samr.gov.cn/wljys/htjg/htsfwb/>，2022 年 12 月 9 日訪問。

在實踐應用層面，這種方法是目前效果最好的，但是缺點也同樣明顯，即其需要大量的人力完成標注工作，可謂是費力費時；而無監督關係抽取就是要把相同關係的範本整合起來，其不需要人工標注的資料，而且可以實現自動提取實體關係。但缺點也同樣明顯，即是準確性不高。因此，有專家提出利用知識庫回標文本自動獲取大量的弱監督資料，這樣就能同時實現關係抽取具有較高準確性與節省人力成本的目的。另外，基於規則的關係抽取方法會先讓語言學專家根據關係抽取的任務設計出一些包含語義特徵、語法與詞彙的模式。之後，在語料文本分析過程中尋找與這些模式相匹配的實例，從而推導出實體之間的語義關係，例如，“合法合規”、“不合法”等；

三是屬性抽取。實體本身自帶的內部特徵用屬性體現，而屬性借助詞性得以體現。不同詞性的詞語在表達合同內容時的重要性也不盡相同，也正因為如此，對分詞和詞性進行標注顯得尤為重要。本文建議通過 NLPPIR 大資料語義智慧分析平臺進行標注，其針對大資料資訊處理的需要，融合了網路搜索技術、自然語言理解及文本挖掘等 13 項功能，將每個詞語按照其詞性分為名詞、動詞及形容詞等不同的類別。

（三）要素加工與關係連結

在對相關知識進行梳理的基礎上，構建知識圖譜的第三步是識別基本法律關係中的要素，並進行拆解、梳理、提取、分類，分析並連結其間的關聯關係，轉化為用節點和邊表述的陳述句。知識圖譜中最基礎的部分就是實體要素，實體之間的關聯分析也是基於其上；關係連結則是將實體節點納入知識圖譜的必要環節，通過對實體間關係的分析，可以整合動態變化的知識，獲得指引和貫通資訊傳遞的路徑。

在合同審查流程中，實體以不同的要素形式程式。在合同審查的主要部分——合同文本和條款中，實體要素為關鍵的專業術語、事物名稱，以及具體法律法規、國標行規中的名詞或名詞短語。除了有關合同內容、格式的審查外，合同審查還包括對於合同簽約主體、簽約程式是否符合當事人約定、是否合法有效等作出判斷。在合同主體的審查過程中，實體要素主要與主體在資質、信用、技術方面的認證有關，該部分資訊可從其他公權力機關開具的證明文件中獲得；在合同締約程式的審查過程中，實體要素主要是關於流程性審查環節的表述，此部分在判斷合同效力的過程中僅起輔助、補充的作用。知識圖譜的構建還取決於分析、搭建實體節點間聯繫的能力。在知識圖譜上連結成更多的“邊”，能夠清楚地將關聯性、流程性環節對應到知識圖譜的具體方面，使得圖譜更能形象直觀地展示資訊間的邏輯。

具體而言，在合同審查方面，通過梳理與該合同相關的法律規範、指導性案例，以法學通說作為解決合同爭議的補充，對知識圖譜所需內容進行整理；接著，在拆分合同審查整體

流程的基礎上提取合同關鍵資訊作為實體節點，並對合同審查不同環節的要素實體進行加工，搭建其間的關係，以此構建知識圖譜的雛形。進一步地，司法人工智慧學習演算法、文本分析和自然語言處理技術，基於知識圖譜存儲的法律專業術語和合同語料庫，實現語料學習、處理與積累，從而構建適用於合同審查的語言模型，與預設的審查標準進行比對，審閱合同中存在的風險和偏差。

（四）知識圖譜存儲

知識存儲要研究的主要問題是如何將上述處理好的知識資料高效存儲。一般而言，知識圖譜存儲主要有兩種方式，即圖資料庫（Graph Database）和 RDF 格式存儲。其中，RDF 格式主要針對資料模型利用 XML 語法進行編碼，以三元組 SPO（subject, predicate, object）的形式存儲資料。目前，業界較主流的 Freebase 知識圖譜，就是採取了 RDF 格式逐行存儲三元組資料。但是依靠 RDF 格式存儲的資料，搜索效率不高。實踐應用過程中，通常採用建立索引的方式提升搜索效率；而圖資料庫則以圖為資料結構存儲資料，其應用程度比 RDF 存儲更加廣泛，代表性的圖資料庫有 Neo4j、JanusGraph 及 HugeGraph 等，而在實踐應用過程中，應用最廣泛的則是 Neo4j。商事活動對效率的要求往往很高，而合同審查的準確性與效率將是企業重點考量的內容。因此，本文建議應用 Neo4j 以完成合同審查知識圖譜存儲工作。

五、知識圖譜在合同審查中的應用：以網路交易合同為例

如上所述，在知識圖譜構建過程中，實體、實體屬性以及他們之間的法律關係是智慧合約系統構建的主要內容。先搭建出一個基於知識圖譜的智慧合約管理系統，這個系統可以有效降低合同成本。其次，在此基礎上利用語義技術，為雲服務和相應軟體提供合同模型。這個系統能夠為提供者和使用者監控合同的執行情況，並跟蹤在此期間各方的義務履行情況。最後是對該系統中合同的合規性進行檢驗，對於潛在的糾紛提供相應的解決機制，最終實現智慧合約系統的有效運轉。合同智慧處理流程如圖 2。



圖 2 合同智慧處理流程圖

在域外，應用司法人工智慧技術進行商業合同文本的審查已有豐富的實踐經驗，如摩根大通研發出來的一款針對商業貸款合同的審查系統，能夠在短短幾秒鐘之內進行曾經需要用

36 萬小時才能完成的工作量。³⁰ 通過梳理結構化的合同文本內容，以其中清晰的要素、明確的規則、標準的術語繪製知識圖譜，在大量而廣泛的樣本和資料之基礎上進行擴展，並輔之以指導性案例（甚至法學通說）進行補充，可以滿足司法人工智慧在合同審查整體流程和各個環節中的要求。

隨著物聯網的日益使用，網路交易合同廣泛使用，從交易前到交易後都可以通過系統進行追蹤。就交易前的合同內容審查方面，通過自動提取合同內容的關鍵資訊，梳理分析合同的各項要素，構建以買賣雙方（特別是供應商、經銷商）、貨品單價、交易數額、交易日期等為節點的實體，以這些實體間的關係為連結的“邊”，構建初步的知識圖譜，進行合同內容一致性的審查，並自動生成對比結果報告。以知識圖譜所存儲的使用者合同的完整視圖，結合其所習得的概念和資料，匯總統計出一個值，如確定網路交易合同的總金額，瞭解其商品庫存量與出售量，或是比較賣方在不同地區幾筆交易之間的價格區別。其次，對於合同主體和程式的審查貫穿于合同風險性審查過程中。利用知識圖譜自動識別風險的能力，定制風險審核項，挖掘可能存在風險的法律關係。合同風險審計分析模型的定制恰恰是知識圖譜在法律人工智慧技術方面的應用之一，在這個模型中，將知識圖譜介面與查詢語言相結合，並對知識圖譜所呈現出來的各項關係進行分析，如供應商表、合同表、客戶名單表等，抓取相應的資料同時進行篩選、整合及解析的工作，由此得到可能存在風險的多級合同列表，以此識別供應商是否存在可能欺詐的行為。而交易中，貨物的裝載、運輸和交付也可以被自動跟蹤。潛在的問題可以立即標記，負責人提醒風險，即可以在風險升級之前立即看到預防選項。最後，即使交易完成，智慧合約系統也繼續作用。系統提供了記錄，可以促進業務流程的精細化程度。對合同進行歸檔管理，將整個流程的資料生成歷史資訊，使得這些資料能夠較為容易地被挖掘、聚合和分析，以備各種查詢。

實際上，智慧合約的控制性應用涉及代碼技術與法律規制之間的互動關係。智慧合約系統將使所有合同成為系統可識別的，這個優勢是巨大的，雖然在當前，要實現這些潛在的效益仍需要真正的人工參與。知識圖譜可以提供大量可利用的、靈活配置的資訊，以說明解決問題。事實上，任何嵌入的參數都能夠實現自動更新並呈現為可識別的。智慧合約對法律的

30 [美] 埃裡克·布林約爾松、[美] 安德魯·麥卡菲：《人工智慧概覽》，載《哈佛商業評論中文版》2017 年第 10 期。

需要源於二者皆是建立在信任機制基礎上，並且同意³¹必須是在知情的情況下作出的。在過去的幾年中，關於符合通用資料條例（GDPR）的資料共用的主要研究工作主要集中在知情同意原則。³² 在企業對企業和企業對消費者資料共用的情況下，當同意的程度需求不夠時，許多中小型企業仍然依賴合同。智慧合約未來的發展前景仍是巨大，無論是智慧合約的糾紛解決機制抑或是救濟機制仍待進一步完善。

六、結語

司法人工智慧是 Web 3.0 時代的司法改革風向，而知識圖譜是司法人工智慧運作的前提和基礎，為彌補現有知識圖譜存在的不足，提升弱人工智慧的司法應用效果，筆者構建了合同審查的知識圖譜，針對合同主體、內容、程式等多方面，對整體流程和各個環節進行風險性審查與合法性評估。體系構建、知識抽取、知識融合和知識存儲是知識圖譜構建的四個過程，基於對資料存儲後搜索效率的考慮，本文建議利用 Neo4j 建立合同審查知識圖譜。在討論構建一般性合同知識圖譜的基礎上，本文以網路交易合同為例展現了合同知識圖譜的具體構建方法。一般性合同審查知識圖譜的構建流程，以整理法律規範和專業知識為第一步，輔之以指導性案例和法學通說；再者從合同中提取實體、加工要素，分析實體間法律關係並連結成“邊”。構建網路交易合同審查的具體知識圖譜，需要在此基礎上，進一步設置對應的審核標準，定制合同風險審核模型。基於合同審查的知識圖譜構建，智慧合約系統等應用場景得以具體化、精細化、模式化，這只是合同審查知識圖譜應用的一個側面，但也讓人看到了知識圖譜應用的無限可能與比較優勢甚至是必然趨勢。目前合同審查的知識圖譜構建仍處於起步階段，但仍然可以期待它將為未來司法的數智化改革賦能。

31 歐盟《通用資料保護條例》（General Data Protection Regulation，以下簡稱 GDPR），第 4 條第 11 款將“同意”定義為：“資料主體的同意是指，資料主體依照其意願自由作出的、特定的、知情的、明確的指示。通過以聲明或清晰肯定的行為作出的該等指示，資料主體表明其同意處理與其相關的個人資料。”

32 See Tauqeer, A., Kurteva, A., Chhetri, T. R., Ahmeti, A., & Fensel, A., Automated GDPR Contract Compliance Verification Using Knowledge Graphs, Information, 447 (2022).