基于要件事实抽取与推理引导的类案检索方法

谢婷婷[1],刘 航[1],禄 蕾[1],李茜梅[1],王凌志[2]

(1. 贵阳市烟草专卖局,贵州 贵阳 510000;2. 贵州大学 计算机科学与技术学院, 贵州 贵阳 550025)

摘 要: 类案检索是提升司法裁判一致性与法律适用精准性的关键技术,在智慧司法体系中日益重要。现有方法多侧重于文本语义相似性与表层匹配,难以建模案件背后的裁判逻辑与法律要件关系。为此,提出一种基于大语言模型要件事实抽取与推理引导的类案检索方法。首先,利用大语言模型与句法分析识别裁判文书中的核心行为事实,构建法律要件事实;其次,引入法律知识图谱进行语义增强与推理引导,提升裁判逻辑的一致性建模能力。实验结果表明,所提方法在Precision@5/10(49.09%、48.93%)、MAP(50.38%)及NDCG@10/20/30(79.27%、84.16%、88.94%)等指标上优于基线模型,在类案匹配中的优势明显,为构建具备裁判逻辑感知能力的类案检索提供了理论支撑与技术路径。

关键词: 智慧司法;类案检索;大语言模型;法律要件;法律知识图谱

DOI: 10. 11907/rjdk. 251542 扫描二维码阅读全文:

中图分类号: TP391.1 **文献标识码:** A **文章编号:** 1672-7800(XXXX)0XX-0001-06

Case Retrieval Method via Legal Element Extraction and Inference Guidance

XIE Tingting[1], LIU Hang[1], LU Lei[1], LI Ximei[1], WANG Lingzhi[2]

(1. *Guiyang Tobacco Monopoly Bureau*, *Guiyang* 510000, *China*;
2. *College of Computer Science and Technology*, *Guizhou University*, *Guiyang* 50025, *China*)

Abstract: Case retrieval is a key technology for improving the consistency of judicial rulings and the accuracy of legal application, and is increasingly important in the smart judicial system. Existing methods often focus on text semantic similarity and surface matching, making it difficult to model the judicial logic and legal requirements behind the case. Therefore, a case retrieval method based on large language model element fact extraction and inference guidance is proposed. Firstly, utilizing the big language model and syntactic analysis to identify the core behavioral facts in the judgment document, and constructing legal element facts; Secondly, introducing a legal knowledge graph for semantic enhancement and reasoning guidance enhances the consistency modeling ability of judicial logic. The experimental results indicate that the proposed method is effective in Precision@5 /10 (49.09%, 48.93%), MAP (50.38%), and NDCG@10 /On indicators such as 20/30 (79.27%, 84.16%, 88.94%), it outperforms the baseline model and has a significant advantage in case matching, providing theoretical support and technical path for building case retrieval with the ability to perceive judicial logic.

Key Words: intelligent judiciary; case retrieval; large language models; legal elements; legal knowledge graphs

0 引言

随着人工智能的快速发展,智慧司法建设不断提速,成为了我国推进司法现代化、实现公平正义的重要战略方向。通过人工智能技术提升审判效率与质量,可实现司法过程的智能化、透明化与标准化。在这一背景下,各类智能司法应用不断涌现,发挥了关键作用[1-4]。

在司法实践中,类案检索作为辅助法官裁判、保障司法公正的重要手段,正逐渐成为法律领域关注的重点[5-7]。随着司法案件数量激增,如何从庞大的裁判文书库中精准、有效地检索出与当前案件在法律适用和裁判逻辑上高

收稿日期: 2025-08-12

基金项目: 贵州省烟草公司贵阳市公司项目(2024)

作者简介: 谢婷婷(1995-),女,硕士,贵阳市烟草专卖局经济师,研究方向为民商法;刘航(1990-),女,硕士,贵阳市烟草专卖局物流师,研究方向为合规管理;禄蕾(1986-),女,贵阳市烟草专卖局政工师,研究方向为民法;李茜梅(1991-),女,贵阳市烟草专卖局政工师,研究方向为民法、行政法;王凌志(2001-),男,贵州大学计算机科学与技术学院硕士研究生,研究方向为自然语言处理。

度一致的先例，成为了法律+AI领域亟待解决的关键问题。传统类案检索方法在逻辑上与文本信息检索别无二致，主要依赖基于关键词的文本匹配或相似度计算，虽在一定程度上实现了自动化检索，但未充分理解法律文本的深层语义，尤其在复杂的法律推理和裁判逻辑层面[8]。因此，如何通过更智能化的方法提升类案检索的精度与智能化水平，仍是司法领域面临的巨大挑战。

近年来，随着大语言模型（ChatGPT、DeepSeek等）和领域知识图谱等技术的兴起，类案检索研究进入了一个新阶段。这些技术虽能精准理解法律语义和案件背景，在文本语义理解上表现优异，但在处理复杂的法律关系和多维度法律信息时性能较弱，存在如何准确捕捉案件要素、如何增强法律推理能力、如何构建高效的模型等问题[9]。

为此，本文旨在提出一种创新的类案检索方法，结合大语言模型的自然语言处理能力与法律要素分析，在精准度和智能化水平上实现突破。具体为，构建一个基于法律要件事实分析的深层语义模型，结合外部法律知识图谱提升检索准确性、推理能力，为法官提供更可靠的决策支持，推动司法智能化进程。主要贡献为：

（1）提出基于要件事实推理分析的类案检索方法。基于句法依存分析与实体识别，结合大语言模型对裁判文书中的主体行为及其相关事实进行结构化抽取，有效构建了时序化、逻辑化的要件事实链条，提升了模型对案件裁判核心信息的聚焦能力。

（2）引入法律知识图谱实现要件知识增强机制。首先，利用多维度法律语义对已抽取的要件事实进行领域知识注入；其次，利用多头注意力机制融合增强案件向量表示，提升模型对法律要件推理过程的理解与建模能力。

1 相关工作

类案检索的发展经历了从传统的统计方法，到基于神经网络与结构化信息融合的复杂模型，再到近年来大语言模型驱动的新阶段。大语言模型凭借其卓越的语义理解与推理能力，为类案检索带来了质的飞跃，逐渐成为了研究的主流方向。

早期，类案检索方法主要基于统计特征，例如TF-IDF、BM25等通过计算词频等文档固有的属性实现文档表示，以进行自动化检索，但缺乏对文本语义的深刻理解[10,11]。随着深度学习兴起，研究者逐步将神经网络应用于案件向量编码。Tran等[12]使用多层感知机（Multilayer Perceptron，MLP）进行案件向量化。Kim等[13]提出Text-CNN方法，在文本领域中通过卷积神经网络获取文本特征。Nigam等[14]结合BM25与神经网络提升了模型在检索空间中的表示能力。然而，这些方法未深入理解法律文档的领域特征。

近年来，基于预训练语言模型的检索方法取得了显著进展，例如BERT、RoBERTa相较于浅层网络，能获取文本更深层次的语义[15,16]。Lawformer、BERT-PLI等模型在处理长文本和建模法律语义方面表现突出[17,18]。Su等[19]提出一种无监督的预训练框架Caseformer，通过自监督学习任务捕捉相似要素，显著提升了模型在低资源条件下的检索性能。然而，这些方法侧重于语义匹配或结构建模，未深入解决案件要件事实和裁判逻辑一致性问题。

随着对法律文书结构性理解的加强，许多研究开始尝试将文档内部结构与外部法律知识相结合，以进一步提升检索效果。Ma等[20]提出一种结构化类案检索框架，利用法律文书的内在结构，避免因文档过长导致信息丢失。Li等[21]将结构信息引入预训练阶段，使SAILER模型能综合分析裁判文书各部分的语义差别。CFGL-LCR是一种基于图神经网络的反事实图学习框架，通过对案件文书进行图形建模，以强化案件元素之间的语义关联，同时通过反事实数据生成缓解训练数据稀缺的问题[22]。CaseLink引入全局案件图捕捉案件之间的语义和法律指控关系，可进一步提升检索性能[23]。虽然，这些方法有效整合了结构化信息，但侧重于语义匹配和结构建模，缺失裁判底层逻辑等更深层次关系。

大语言模型也被应用于类案检索任务。Deng等[24]提出KELLER框架，通过领域知识微调的大语言模型对案件文本进行重构，从而有效过滤冗余信息。Kim等[25]提出LegalSearchLM，将法律案件检索重新定义为法律要素生成任务，采用受约束的解码技术直接生成与查询案件相关的内容，克服了传统嵌入式方法在语义表达上的局限性。然而，上述方法未充分发挥其对于复杂法律文本的理解能力，无法发掘事件中蕴含的裁判逻辑[26]。

相较于现有方法，本文方法将案件要件事实与裁判逻辑对齐，通过“子事实抽取—要件事实分析—知识增强”三阶段框架，结合大语言模型的推理能力和外部法律知识图谱，实现对案件深层次裁判逻辑的理解与表达，从而在类案检索中提供更精准的匹配结果。具体地，以发掘司法判决共性为目标，从法官司法实务中对“要件事实”的关注点出发，围绕“子事实抽取—要件事实分析—要件知识增强”构建案件裁判逻辑的深层次语义表示，并结合大语言模型的推理能力与外部法律知识图谱，精准检索相似案例。

2 本文方法

2.1 模型架构

图1为本文模型框架，由件事实分析（图1(a)）与要件知识增强（图1(b)）两个核心模块组成，旨在解决传统类案检索中存在的语义理解不充分、案件判决共性语义表示不足等问题。具体为，引入子事实识别、要件事实分析及要件推理机制，通过深层次理解与高效比较案件文本，发掘各案件内部蕴含的基于要件的裁判底层逻辑，从而提升

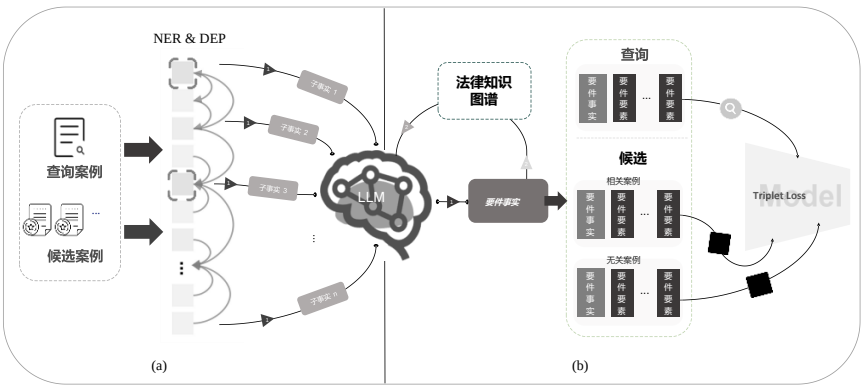

Fig. 1　Model architecture

图1　模型架构

类案检索的准确性与鲁棒性。

2.2　要件事实分析

模型在要件事实分析模块中,通过句法依存分析候选集和查询案例文书中的行为主体S、行为方式V、客体对象O,并利用层级化抽取与语义重构方法得到子事实。

首先,设输入案件文本为$\boldsymbol{T}$,通过命名实体识别与句法依存分析得到每个案件的子事实与子事实集合分别为f、$\boldsymbol{F}$。

$$T \to \{f_1, f_2, ..., f_N\} = F \tag{1}$$

其次,通过LLM对已抽取的子事实集合F进行分析,结合案例文本数据的推理字段(*reason*)识别F中的关键子事实,并依照时序构造案件的要件事实f_{keys_c}。

$$f_{keys\_c} = M(Prompt|reason, F) \tag{2}$$

式中:*Prompt*为要件事实分析的提示词;M为大模型推理操作;f_{keys_c}为候选案例的要件事实。

最后,分析出候选集的要件事实后,采用单样本学习方法(one-shot)替代查询案例缺失的reason字段,以获得查询案例进行要件事实分析f_{keys_q}。

$$f_{keys\_q} = M(Prompt|\left(F_c, f_{keys\_c}\right), F_q) \tag{3}$$

式中:F_c为随机选取的候选案例的子事实集合;F_q为查询案例的子事实集合。

2.3　要件知识增强

在要件知识增强模块中,旨在对前述模块得到的要件事实进行处理,从而获得要件的逻辑信息,强化模型对案例间判决共性的捕捉。

首先,引入法律知识图谱进行知识注入,增强要件事实文本中的领域知识;其次,通过大语言模型推理得到案例的知识增强要素集合$\boldsymbol{C}$。表1为知识增强要素聚焦的4个维度。整个要件知识增强过程为:

$$C = M(Prompt_k|f_{keys}, KG) \tag{4}$$

式中:KG为要件事实中由知识图谱查询得到的实体知识;$Prompt_k$为要件知识增强的提示词。

Table 1　Key dimensions of focus in element analysis

表1　要件分析关注的关键维度

维度	内容	大语言模型提示词
主体状态	当事人行为状态、认知意图等	请从案件文本中抽取并识别各主体的法律身份、程序角色及其在案件中的行为表现与主观认知。要求:①分析须以文本证据或合理法律推理为基础,避免主观臆断;②结果应包含主体名称、诉讼地位(原告/被告/第三人等)、行为状态(如履行/不履行/违法/合规)以及认知意图(故意/过失/无过错/不确定);③如文本未涉及相关要素,应标注为"未提及"
时序关系	行为时间顺序、认定推理顺序等	请从案件文本中抽取关键事实事件,并对其进行时间顺序和逻辑推理顺序的整理。要求:①明确事件的先后、并行或因果关系;②在条件允许时标注绝对时间(日期)或相对时间(如"事发前""次日");③若时间关系无法确定,则标注为"未明确"
风险预见	推理链可能涉及的法律风险	请识别案件事实链条中潜在的法律风险及其可预见性。要求:①区分当事人已明确知晓的风险与依据合理人标准可预见的风险;②为每一风险标注可能的法律后果(如民事赔偿、刑事责任、行政处罚);③如文本未提供足够信息,应标注为"不确定"
判决影响	当事人、司法裁决体系、社会影响	请分析判决或裁决可能带来的影响。要求:①从当事人层面(如经济损失、自由限制)、司法体系层面(如先例价值、规则适用)以及社会层面(如公共利益、行业规范)三个维度进行归纳;②明确各类影响的性质(积极/消极/中性)及程度(高/中/低);③如文本未涉及,应标注为"未提及"

首先,通过大模型要件知识增强得到要素集合$\boldsymbol{C}$,并将其与案件的要件事实f_{keys}采用多头注意力机制进行特征融合,得到表示整个案件的裁判逻辑与核心内容的要件特征$\boldsymbol{X}_k$。

$$\boldsymbol{X}_k = \text{concat}\left(Attention_{1 \to H}\left(Q, K, V\right)\right)\boldsymbol{W}_o + b \tag{5}$$

式中:$Attention_{1 \to H}$为从第1个到第H个头的注意力计算;$\boldsymbol{W}_o$为一个可学习的权重矩阵,以对拼接后的特征进行线性变换;b为偏置项。

其次,Q、K、V的计算公式为:

$$\boldsymbol{Q} = \boldsymbol{C}\boldsymbol{W}_Q, \boldsymbol{K} = f_{keys}\boldsymbol{W}_K, \boldsymbol{V} = f_{keys}\boldsymbol{W}_V \tag{6}$$

式中:$\boldsymbol{W}_Q$、$\boldsymbol{W}_K$和$\boldsymbol{W}_V$为可学习的权重矩阵。

最后,将通过多头注意力机制融合的要件特征X_k用作下游检索模型的训练特征,并将检索任务中广泛应用的三元组损失作为目标函数进行训练。

$$Loss = Max\left(d\left(X_{kq}, X_{kp}\right) - d\left(X_{kq}, X_{kn}\right) + m, 0\right) \quad (7)$$

式中：d 表示向量间距离本方法采用欧式距离；m 表示预设裕度；X_{kq}、X_{kp}、X_{kn} 分别表示查询案例、相关案例、无关案例的特征表示。

3 实验结果与分析

3.1 实验设置

本文实验编程语言为 Python，在 PyCharm Professional 2024.3.3 平台上实现，GPU 为 NVIDIA A100-40GB。同时，在 Ubuntu 20.04 操作系统下，基于 Pytorch 深度学习框架进行模型训练、微调与推理。大语言模型为 Qwen3-14B（https://huggingface.co/Qwen/Qwen3-14B），知识图谱采用 TechKG 法律知识库（https://huggingface.co/Qwen/Qwen3-14B），选用 BGE-base-zh-v1.5 模型（https://huggingface.co/BAAI/bge-base-zh-v1.5）作为下游文本嵌入模型，其余实验通用参数设置如表 2 所示[27,28]。

Table 2 Experimental parameters setting

表 2 实验参数设置

参数	设定值
批次大小（Batchsize）	4
学习率（Learning rate）	2×10^5
权重衰减（Weight decay）	0.01
训练轮次（Epoch）	200
学习率预热（Warmup ratio）	
损失函数（Loss function）	0.1
三元组损失函数（Triplet Loss）	

3.1.1 实验数据集

本文使用“法研杯”（CAIL2022）提供的类案检索赛道数据集 LeCaRD（https://github.com/myx666/LeCaRD）进行实验[29]。该数据集的案例数量为 10 802，每个候选案例包括基本事实、法院推理、判决结果、判决全文等内容，经过分析推理及知识增强后，将其作为模型训练特征，如表 3 所示。

Table 3 LeCARD dataset

表 3 LeCARD 数据集

属性	比例/取值范围
查询/候选案例	1：100
相似度标签取值	{0，1，2，3}
查询案例平均长度	445
候选案例平均长度	6 319
训练集：验证集：测试集	70∶15∶15

3.1.2 性能评价指标

本文选用信息检索领域中被广泛应用的多个指标来衡量系统性能，包括平均精度均值（MAP）、不同 K 值下的精确率（Precision@K）及归一化折损累计增益（NDCG@K）。其中，Precision@K 用于衡量检索结果中前 K 个条目中相关案例所占的比例，强调排序前列结果的准确性；MAP 用于评估整体排序效果的优劣，体现全局性能；NDCG@K 结合相关性等级与位置权重，对前 K 位结果进行加权评价以反映排序质量。具体计算公式为：

$$Precision@K = \frac{1}{K}\sum_{i=1}^{K} y_i \quad (8)$$

$$AP = \frac{1}{m}\sum_{k=1}^{n} P@k \quad (9)$$

$$MAP = \frac{1}{Q}\sum_{q=1}^{Q} AP_q \quad (10)$$

$$DCG@K = \sum_{i=1}^{K}\frac{2^{rel(i)}-1}{\log_2(i+1)} \quad (11)$$

$$IDCG@K = \sum_{i=1}^{\{rel_K\}}\frac{2^{rel(i)}-1}{\log_2(i+1)} \quad (12)$$

$$NDCG@K = \frac{DCG@K}{IDCG@K} \quad (13)$$

式中：y_i 表示候选文档的相似度标签；Q 表示查询案例的总数量；APq 表示查询 q 的平均准确率；$rel(i)$ 为第 i 个候选文档的相关性标签；$\{rel_K\}$ 表示获取前 K 个文档的相似度排序。

3.1.3 基线模型

为了评估本文方法的有效性，选取多个具有代表性的模型作为基线，涵盖从传统方法到近年来表现突出的深度学习模型。

TF-IDF 是一种经典的词频加权方法，计算词项在文档与语料库中的频率以反映其重要性，本文将查询与候选文档分别向量化，通过余弦相似度进行匹配[10]。BM25 是一种基于概率统计的文本检索模型，通过对词项频率、文档长度等因素进行加权计算，以衡量查询与文档之间的相关性[11]。Text-CNN 是一种典型的卷积神经网络文本表示模型，先利用不同窗口大小的卷积核提取局部语义特征，再通过最大池化聚合关键信息[13]。BERT 是一种基于 Transformer Encoder 的双向预训练语言模型，通过掩码语言建模任务获得更丰富的语义表示[15]。RoBERTa 是 Bert 的改进版本，通过更多数据与训练步骤提升语义建模能力[16]。Lawformer 是基于 Longformer 架构优化而来的法律专用预训练语言模型，能有效处理长篇法律文本，捕捉文书中的长距离依赖特征，适用于建模判决书等结构化法律内容[17]。BERT-PLI 是专门针对法律案件匹配场景设计的文本推理模型，通过精细建模查询与候选案例之间的双向交互关系，以提升法律语义表示的判别力[18]。SAILER 是一种面向法律文本匹配任务的多层次预训练模型，融合了案件事实、法条适用、法律推理等维度信息，通过多任务学习提升模型在法律检索任务中的表现[21]。

3.2 实验结果

3.2.1 比较实验

为了全面评估本文方法的有效性，将其与前述基线模型进行比较，实验结果如表 4 所示。由此可知，TF-IDF、BM25 因未有效建模文本深层语义与法律逻辑，在 MAP、Precision@K 和 NDCG@K 等指标上表现不佳；Text-CNN 尽

管能提取部分局部特征,但难以捕捉长文本中的复杂法律关系,整体效果欠佳。相较而言,预训练语言模型BERT与RoBERTa显著提升了类案匹配的语义建模能力,尤其在NDCG指标上体现出了更优的排序能力。专为法律检索任务设计的BERT-PLI与Lawformer进一步增强了对法律语言和裁判逻辑的理解,表现更优。本文方法在所有评价指标上成绩最优,Precision@5、MAP、NDCG@30分别为49.09%、50.38、88.94,相较于SAILER均有小幅度提升,验证了要件事实推理与知识增强机制,能有效提升类案匹配准确性与合理性。

Table 4　Comparative experiment results

表4　比较实验结果　　**(%)**

方法	Precision		MAP	NDCG		
	@5	@10		@10	@20	@30
BM25	38.31	39.90	27.31	62.40	72.64	76.72
Text-CNN	26.24	25.99	20.65	54.07	58.41	64.38
BERT	33.18	32.36	27.29	60.40	64.04	68.32
RoBERTa	35.21	34.71	33.74	63.13	67.54	72.77
Lawformer	42.33	41.52	45.94	72.69	80.16	79.00
BERT-PLI	47.36	46.47	50.2	74.46	79.38	84.17
SAILER	46.25	47.41	48.96	74.30	78.65	82.16
Ours	**49.09**	**48.93**	**50.38**	**79.27**	**84.16**	**88.94**

3.2.2　鲁棒性实验

为了进一步验证所提方法在不同模型结构下的通用性与鲁棒性,基于若干代表性模型(BM25、Text-CNN、BERT、Lawformer),引入要件事实分析与知识增强机制,以比较模型增强前后的性能变化,实验结果如表5所示(*表示经过本文方法增强后的基线模型)。由此可知,引入要件结构建模与知识注入方法后,所有模型性能均得到了提升。BM25经要件增强后,MAP从27.31提升至36.86;Text-CNN的Precision@10由25.99提升至49.33;BERT与Lawformer分别在多个指标上提升了5%~10%。

综上,本文提出的要件事实驱动机制,不仅适用于结构复杂的预训练语言模型,还能显著提升传统检索方法与浅层网络模型的判例匹配能力,具有良好的跨模型泛化能力与鲁棒性,进一步验证了司法裁判过程中的核心"要件事实"在类案识别中的关键作用,反映出大语言模型在推理与法律知识融合方面优势。

Table 5　Robustness experiment results

表5　鲁棒性实验结果　　**(%)**

方法	Precision		MAP	NDCG		
	@5	@10		@10	@20	@30
BM25	38.31	39.90	27.31	62.40	72.64	76.72
BM25*	**41.17**	**44.82**	**36.86**	**64.99**	**76.97**	**81.19**
Text-CNN	26.24	25.99	20.65	54.07	58.41	64.38
Text-CNN*	**44.87**	**49.33**	**40.74**	**68.14**	**75.80**	**85.29**
BERT	33.18	32.36	27.29	60.40	64.04	68.32
BERT*	**36.12**	**35.15**	**33.30**	**74.38**	**80.28**	**85.03**
Lawformer	42.33	41.52	45.94	72.69	80.16	79.00
Lawformer*	**45.20**	**45.16**	**49.11**	**78.71**	**84.32**	**87.51**

4　结语

本文提出一种融合要件事实抽取与推理引导的类案检索方法,以提升裁判要点识别与语义匹配的精准性。首先,通过大语言模型抽取子事实,构建法律要件结构;其次,结合知识图谱实现语义增强与推理引导,强化了模型对案件争议焦点的建模能力。

在LeCARD数据集上的实验表明,本文方法在Precision@5、MAP、NDCG@30等指标上均优于主流基线模型,验证了要素驱动的类案建模范式的有效性与通用性。同时,实验结果证明了语言模型与法律知识融合的应用潜力,为构建具备法律逻辑感知能力的智能类案检索系统提供了可行路径。

未来,尝试将模型拓展至多类型文书处理与跨域推理等场景,以推动司法AI向高可信、高可解释方向发展。

参考文献:

[1] LIU P X. China-style innovation of intelligent justice[J]. Journal of National Prosecutors College, 2021, 29(3): 81-101.
刘品新. 智慧司法的中国创新[J]. 国家检察官学院学报, 2021, 29(3): 81-101.

[2] CHEN L L. On construction of digital community in court: based on artificial intelligence-assisted judicial adjudication[J]. Law Science, 2024(1): 127-140.
陈罗兰. 论法院数字共同体的构建:以人工智能辅助司法为视角[J]. 法学, 2024(1): 127-140.

[3] JIAO B Q, ZHAO Y. On the impact of artificial intelligence on judges' thinking[J]. Seeking Truth, 2022, 49(4): 115-125.
焦宝乾, 赵岩. 人工智能对法官思维的影响[J]. 求是学刊, 2022, 49(4): 115-125.

[4] XU J F, SUN F H, CHEN Q W, et al. A new pattern framework and innovative practices in the smart court system-of-systems engineering project of China[J]. Strategic Study of CAE, 2022, 24(4): 105-120.
许建峰, 孙福辉, 陈奇伟, 等. 我国智慧法院体系工程的模式框架和创新实践[J]. 中国工程科学, 2022, 24(4): 105-120.

[5] DENG W P, ZHAO W Y. Case retrieval: a technical approach to judicial governance in the digital era[J]. Nomocracy Forum, 2022(4): 37-49.
邓伟平, 赵文宇. 类案检索:数字时代司法治理的技术进路[J]. 法治论坛, 2022(4): 37-49.

[6] SUN Y. The judicial application and improvement of case retrieval[J]. Legal Method, 2022, 39(2): 268-281.
孙跃. 类案检索的司法适用及其完善[J]. 法律方法, 2022, 39(2): 268-281.

[7] WANG S Z. Intelligence of similar cases retrieval in the perspective of normative guidance[J]. Journal of Law Application, 2021(9): 150-159.
王肃之. 规范指导视域下类案检索的智慧化[J]. 法律适用, 2021(9): 150-159.

[8] QIAN Y G, QIN X L, ZHANG S Q, et al. Survey of texts semantic matching based on deep learning[J]. Software Guide, 2022, 21(12): 252-261.
钱杨舸, 秦小林, 张思齐, 等. 基于深度学习的文本语义匹配综述[J]. 软件导刊, 2022, 21(12): 252-261.

[9] FENG Y, LI C, NG V. Legal case retrieval: a survey of the state of the art [C]// Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics, 2024: 6472-6485.

[10] SALTON G, BUCKLEY C. Term-weighting approaches in automatic text retrieval [J]. Information processing & management, 1988, 24(5): 513-523.

[11] ROBERTSON S E, WALKER S, JONES S, et al. Okapi at TREC-3 [C]// Third Text Retrieval Conference, 1994: 109-126.

[12] TRAN V, NGUYEN M L, SATOH K. Building legal case retrieval systems with lexical matching and summarization using a pre-trained phrase scoring model [C]// 17th International Conference on Artificial Intelligence and Law, 2019: 275-282.

[13] KIM Y. Convolutional neural networks for sentence classification [C]// 2014 Conference on Empirical Methods in Natural Language Processing, 2014: 1746-1751.

[14] NIGAM S K, GOEL N, BHATTACHARYA A. Nigam@COLIEE-22: legal case retrieval and entailment using cascading of lexical and semantic-based models [C]// JSAI International Symposium on Artificial Intelligence, 2022: 96-108.

[15] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [C]// 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019: 4171-4186.

[16] LIU Y, OTT M, GOYAL N, et al. Roberta: a robustly optimized Bert pretraining approach [DB/OL]. https://arxiv.org/abs/1907.11692.

[17] XIAO C, HU X, LIU Z, et al. Lawformer: a pre-trained language model for chinese legal long documents [J]. AI Open, 2021, 2: 79-84.

[18] SHAO Y, MAO J, LIU Y, et al. BERT-PLI: modeling paragraph-level interactions for legal case retrieval [C]// 29th International Joint Conference on Artificial Intelligence, 2020: 3501-3507.

[19] SU W, AI Q, WU Y, et al. Pre-training for legal case retrieval based on inter-case distinctions [J]. ACM Transactions on Information Systems, 2025, 43(5): 1-27.

[20] MA Y, WU Y, AI Q, et al. Incorporating structural information into legal case retrieval [J]. ACM Transactions on Information Systems, 2023, 42(2): 1-28.

[21] LI H, AI Q, CHEN J, et al. SAILER: structure-aware pre-trained language model for legal case retrieval [C]// Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2023: 1035-1044.

[22] ZHANG K, CHEN C, WANG Y, et al. CFGL-LCR: a counterfactual graph learning framework for legal case retrieval [C]// 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023: 3332-3341.

[23] TANG Y, QIU R, YIN H, et al. Caselink: Inductive graph learning for legal case retrieval [C]// 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2024: 2199-2209.

[24] DENG C, MAO K, DOU Z. Learning interpretable legal case retrieval via knowledge-guided case reformulation [DB/OL]. https://arxiv.org/abs/2406.19760.

[25] KIM C, LEE J, HWANG W. LegalSearchLM: rethinking legal case retrieval as legal elements generation [DB/OL]. https://arxiv.org/abs/2505.23832.

[26] YANG A, LI A, YANG B, et al. Qwen3 technical report [DB/OL]. https://arxiv.org/abs/2505.09388.

[27] REN F, HOU Y, LI Y, et al. Techkg: a large-scale chinese technology-oriented knowledge graph [DB/OL]. https://arxiv.org/abs/1812.06722.

[28] XIAO S, LIU Z, ZHANG P, et al. C-pack: packed resources for general chinese embeddings [DB/OL]. https://arxiv.org/abs/2309.07597.

[29] MA Y, SHAO Y, WU Y, et al. LeCaRD: a legal case retrieval dataset for Chinese law system [C]// 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021: 2342-2348.

（责任编辑：刘嘉文）