(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 120317230 A
(43)申请公布日 2025.07.15

(21)申请号 202510470764.5
(22)申请日 2025.04.15
(71)申请人 东莞市元诉科技有限公司
地址 523000 广东省东莞市南城街道莞太路南城段111号民间金融大厦1栋2606室之06
(72)发明人 陈欢 陆志贤 周海璇
(74)专利代理机构 广东居安专利商标代理事务所(普通合伙) 44995
专利代理师 刘琼
(51)Int.Cl.
G06F 40/186(2020.01)
G06F 40/205(2020.01)
G06Q 50/18(2012.01)
G06N 3/0455(2023.01)
G06F 16/35(2025.01)

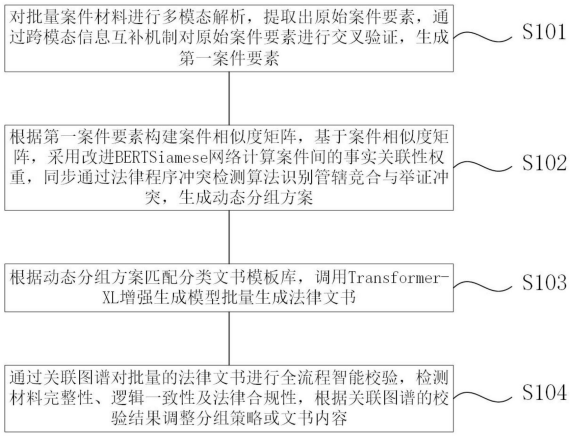
权利要求书3页 说明书17页 附图2页

(54)发明名称

一种用于司法批量立案的智能化处理方法及系统

(57)摘要

本发明公开了一种用于司法批量立案的智能化处理方法及系统,所述方法具体包括:通过跨模态信息互补机制对原始案件要素进行交叉验证,生成第一案件要素;根据第一案件要素构建案件相似度矩阵,基于案件相似度矩阵,采用改进BERT-Siamese网络计算案件间的事实关联性权重,同步通过法律程序冲突检测算法识别管辖竞合与举证冲突,生成动态分组方案;根据动态分组方案匹配分类文书模板库,调用Transformer-XL增强生成模型批量生成法律文书;通过关联图谱对批量的法律文书进行全流程智能校验,检测材料完整性、逻辑一致性及法律合规性。本发明实现了司法批量立案的智能化、高效化和精准化,提高了案件处理效率和文书生成的质量。



CN 120317230 A

1.一种用于司法批量立案的智能化处理方法,其特征在于,所述方法具体包括:

对批量案件材料进行多模态解析,提取出原始案件要素,通过跨模态信息互补机制对原始案件要素进行交叉验证,生成第一案件要素;

根据第一案件要素构建案件相似度矩阵,基于案件相似度矩阵,采用改进BERT-Siamese网络计算案件间的事实关联性权重,同步通过法律程序冲突检测算法识别管辖竞合与举证冲突,生成动态分组方案;

根据动态分组方案匹配分类文书模板库,调用Transformer-XL增强生成模型批量生成法律文书;

通过关联图谱对批量的法律文书进行全流程智能校验,检测材料完整性、逻辑一致性及法律合规性,根据关联图谱的校验结果调整分组策略或文书内容。

2.根据权利要求1所述的方法,其特征在于,所述对批量案件材料进行多模态解析,具体包括:

采用预构建的法律NLP模型对批量案件材料中的起诉状、证据清单进行实体识别与关系抽取,输出文本特征向量集;

通过改进OCR-ResNet模型提取批量案件材料中的表格、签章及物证链,生成图像特征矩阵;

使用ASR-Transformer模型将批量案件材料中的庭审录音转换为文本,提取出包含争议焦点的语音时序特征。

3.根据权利要求2所述的方法,其特征在于,所述通过跨模态信息互补机制对原始案件要素进行交叉验证,生成第一案件要素,具体包括:

将原始案件要素按照要素类型分为事实要素、法律要素和程序要素;

为每个原始案件要素分别添加来源模态标识,所述来源模态标识用于标识其来源于文本、图像或语音;

基于来源模态标识,计算每个原始案件要素中涉及的文本特征向量集、图像特征矩阵和语音时序特征之间的余弦相似度,通过可学习权重参数动态调节文本特征向量集、图像特征矩阵和语音时序特征之间的模态贡献比例;

当所述余弦相似度低于预设相似度阈值时,触发融合修正操作,将文本特征向量集、图像特征矩阵和语音时序特征进行加权融合,生成跨模态要素表示;

基于要素类型和跨模态要素表示,将原始案件要素转换为第一案件要素。

4.根据权利要求3所述的方法,其特征在于,所述根据第一案件要素构建案件相似度矩阵,具体包括:

根据文本特征向量集、图像特征矩阵和语音时序特征之间的余弦相似度及对应的模态贡献比例,计算获得跨模态验证置信度;

基于跨模态验证置信度的均值,分别计算事实要素、法律要素及程序要素的动态权重,形成案件相似度矩阵。

5.根据权利要求1所述的方法,其特征在于,所述基于案件相似度矩阵,采用改进BERT-Siamese网络计算案件间的事实关联性权重,同步通过法律程序冲突检测算法识别管辖竞合与举证冲突,生成动态分组方案,具体包括:

将案件相似度矩阵输入至改进BERT-Siamese网络,通过双塔结构对案件要素对进行编

码,获得编码案件特征;

对编码案件特征进行多维权重融合,生成融合特征表示,采用混合相似度度量方法计算融合特征表示中的案件间的事实关联性权重;

基于案件相似度矩阵,通过比较案件管辖法院是否一致,并结合案件当事人集合的交集判断冲突程度,获得第一冲突检测结果;

基于案件相似度矩阵,通过比较案件举证期限的时间重叠度,并结合证据链相似度综合评估冲突强度,获得第二冲突检测结果;

根据事实关联性权重、第一冲突检测结果和第二冲突检测结果,采用改进谱聚类算法生成动态分组方案。

6.根据权利要求1所述的方法,其特征在于,所述根据动态分组方案匹配分类文书模板库,调用Transformer-XL增强生成模型批量生成法律文书,具体包括:

通过计算动态分组方案中各案件组的特征向量与分类文书模板库的每个模板的特征锚点之间的加权余弦相似度,确定法律文书模板,所述特征锚点包括案件类型编码、标的额分段编码和程序阶段编码;

构建Transformer-XL增强生成模型,所述Transformer-XL增强生成模型包括案件组级记忆单元和双通道生成单元,所述案件组级记忆单元用于通过记忆矩阵缓存同组案件的共性特征,并将记忆矩阵与当前案件隐藏状态拼接形成增强特征表示,所述双通道生成单元用于输入案件组共性特征和个案专属特征,并通过联合概率分布生成法律文书内容;

将法律文书模板输入至Transformer-XL增强生成模型中,批量生成法律文书。

7.根据权利要求1所述的方法,其特征在于,所述通过关联图谱对批量的法律文书进行全流程智能校验,检测材料完整性、逻辑一致性及法律合规性,具体包括:

设置关联图谱,所述关联图谱的顶点集包括案件要素、法律条文、历史判例和法律文书,所述动态关联图谱的边集包括案件要素相似度、任一案件要素与任一引用法条的关系、任一法律文书与任一历史判例之间的逻辑冲突关系;

根据关联图谱批量检测每个法律文书的案件要素与法律文书的映射覆盖率,获得材料完整性校验结果;

根据关联图谱批量检测每个法律文书与每个历史判例之间的逻辑冲突关系,获得逻辑一致性校验结果;

根据关联图谱批量检测每个法律文书的文书内容与每个引用法条之间的匹配度,获得法律合规性校验结果。

8.根据权利要求7所述的方法,其特征在于,所述根据关联图谱的校验结果调整分组策略或文书内容,具体包括:

基于材料完整性校验结果、逻辑一致性校验结果和法律合规性校验结果,确定错误边集合和错误特征模式;

根据错误边集合重构案件相似度矩阵,根据错误特征模式对法律文书的文书内容进行修正处理。

9.一种用于司法批量立案的智能化处理系统,其特征在于,所述系统具体包括:

第一处理模块,用于对批量案件材料进行多模态解析,提取出原始案件要素,通过跨模态信息互补机制对原始案件要素进行交叉验证,生成第一案件要素;

第二处理模块,用于根据第一案件要素构建案件相似度矩阵,基于案件相似度矩阵,采用改进BERT-Siamese网络计算案件间的事实关联性权重,同步通过法律程序冲突检测算法识别管辖竞合与举证冲突,生成动态分组方案;

第三处理模块,用于根据动态分组方案匹配分类文书模板库,调用Transformer-XL增强生成模型批量生成法律文书;

第四处理模块,用于通过关联图谱对批量的法律文书进行全流程智能校验,检测材料完整性、逻辑一致性及法律合规性,根据关联图谱的校验结果调整分组策略或文书内容。

10.一种计算机设备,其特征在于,包括:存储器和处理器及存储在存储器上的计算机程序,当所述计算机程序在处理器上被执行时,实现如权利要求1至8中任一项所述的用于司法批量立案的智能化处理方法。

一种用于司法批量立案的智能化处理方法及系统

技术领域

[0001] 本发明涉及人工智能技术领域,尤其涉及一种用于司法批量立案的智能化处理方法及系统。

背景技术

[0002] 在司法领域,随着案件数量的持续增长和司法资源的相对有限性,如何高效、准确地处理批量案件成为司法系统面临的重要挑战。传统的司法批量立案方法主要依赖于人工操作和预设规则,存在诸多局限性,难以满足现代司法实践的需求。

[0003] 首先,传统方法在案件合并时,主要依赖人工预设的简单规则进行分组。这种分组方式缺乏对案件要素深度语义理解的动态调整能力,无法根据案件要素的复杂关联性进行智能分组。因此,在面对复杂案件时,往往导致分组不合理,影响后续处理效率。

[0004] 其次,传统法律文书生成方式主要基于固定模板的填充式生成,即根据预设的文书模板,将案件信息手动或简单替换填充至模板中。这种方式无法根据案件特征自动生成个性化法律文书,难以体现案件的独特性和法律适用的灵活性。随着案件复杂性的增加,这种生成方式的局限性愈发明显。

[0005] 最后,传统司法批量立案流程中,材料审核、要素提取、文书生成等环节往往相互独立,缺乏智能协同机制。数据流转主要依赖人工干预,不仅效率低下,而且容易出错。各环节之间的信息孤岛现象严重,无法实现信息的有效共享和利用。

[0006] 此外,传统方法在处理案件时,虽然也会参考法律条文和历史判例,但这种利用往往停留在表面匹配阶段。缺乏动态更新的法律知识图谱支持,无法深入挖掘法律条文与案件事实之间的复杂关联性。因此,在法律适用和文书生成过程中,往往难以做到精准、高效。

发明内容

[0007] 本发明的目的在于提供一种用于司法批量立案的智能化处理方法及系统,实现了司法批量立案的智能化、高效化和精准化,提高了案件处理效率和文书生成的质量,减少了人工干预,降低了错误率,以解决上述现有技术问题的至少之一。

[0008] 第一方面,本发明提供了一种用于司法批量立案的智能化处理方法,所述方法具体包括:

[0009] 对批量案件材料进行多模态解析,提取出原始案件要素,通过跨模态信息互补机制对原始案件要素进行交叉验证,生成第一案件要素;

[0010] 根据第一案件要素构建案件相似度矩阵,基于案件相似度矩阵,采用改进BERT-Siamese网络计算案件间的事实关联性权重,同步通过法律程序冲突检测算法识别管辖竞合与举证冲突,生成动态分组方案;

[0011] 根据动态分组方案匹配分类文书模板库,调用Transformer-XL增强生成模型批量生成法律文书;

[0012] 通过关联图谱对批量的法律文书进行全流程智能校验,检测材料完整性、逻辑一

致性及法律合规性,根据关联图谱的校验结果调整分组策略或文书内容。

[0013] 第二方面,本发明提供了一种用于司法批量立案的智能化处理系统,所述系统具体包括:

[0014] 第一处理模块,用于对批量案件材料进行多模态解析,提取出原始案件要素,通过跨模态信息互补机制对原始案件要素进行交叉验证,生成第一案件要素;

[0015] 第二处理模块,用于根据第一案件要素构建案件相似度矩阵,基于案件相似度矩阵,采用改进BERT-Siamese网络计算案件间的事实关联性权重,同步通过法律程序冲突检测算法识别管辖竞合与举证冲突,生成动态分组方案;

[0016] 第三处理模块,用于根据动态分组方案匹配分类文书模板库,调用Transformer-XL增强生成模型批量生成法律文书;

[0017] 第四处理模块,用于通过关联图谱对批量的法律文书进行全流程智能校验,检测材料完整性、逻辑一致性及法律合规性,根据关联图谱的校验结果调整分组策略或文书内容。

[0018] 第三方面,本发明提供了一种计算机设备,包括:存储器和处理器及存储在存储器上的计算机程序,当所述计算机程序在处理器上被执行时,实现如上述方法中任一项所述的用于司法批量立案的智能化处理方法。

[0019] 第四方面,本发明提供了一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器运行时,实现如上述方法中任一项所述的用于司法批量立案的智能化处理方法。

[0020] 与现有技术相比,本发明具有以下技术效果的至少之一:

[0021] 1、本发明实现了司法批量立案的智能化、高效化和精准化,提高了案件处理效率和文书生成的质量,减少了人工干预,降低了错误率。

[0022] 2、本发明采用法律NLP模型、改进OCR-ResNet模型和ASR-Transformer模型,对批量案件材料中的文本、图像和语音信息进行全面解析,提取出丰富的案件要素特征,为后续的跨模态信息互补和案件相似度计算提供了坚实基础,提高了要素提取的准确性和全面性。

[0023] 3、本发明通过计算不同模态特征之间的余弦相似度,并动态调节模态贡献比例,实现了对原始案件要素的交叉验证。当相似度低于预设阈值时,触发融合修正操作,生成更准确的跨模态要素表示,有效提升了案件要素的准确性和可靠性,为后续的案件相似度计算和动态分组提供了有力支持。

[0024] 4、本发明基于跨模态验证置信度,计算事实要素、法律要素及程序要素的动态权重,形成案件相似度矩阵,为后续的案件分组和文书生成提供了科学依据,使得分组更加合理,文书生成更加精准。

[0025] 5、本发明通过改进BERT-Siamese网络和法律程序冲突检测算法,实现了对案件间事实关联性权重的准确计算,并同步识别出管辖竞合与举证冲突,为生成动态分组方案提供了有力支持,使得分组更加符合司法实践需求,提高了案件处理的效率和准确性。

[0026] 6、本发明通过计算动态分组方案中各案件组的特征向量与分类文书模板库的相似度,确定最合适的法律文书模板,并调用Transformer-XL增强生成模型批量生成法律文书,实现了法律文书的个性化生成,提高了文书生成的质量和效率,降低了人工成本。

[0027] 7、本发明设置关联图谱,对批量的法律文书进行全流程智能校验,检测材料完整性、逻辑一致性及法律合规性,有效提升了法律文书的质量,减少了法律风险,为司法决策提供了有力保障。

[0028] 8、本发明基于校验结果,确定错误边集合和错误特征模式,重构案件相似度矩阵,并对法律文书的文书内容进行修正处理,实现了对分组策略和文书内容的动态调整,进一步提高了案件处理的准确性和效率。

附图说明

[0029] 为了更清楚地说明本申请实施例中的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本申请的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0030] 图1是本发明一实施例提供的用于司法批量立案的智能化处理方法的流程示意图;

[0031] 图2是本发明一实施例提供的用于司法批量立案的智能化处理系统的结构示意图;

[0032] 图3是本发明一实施例提供的一种计算机设备的结构示意图。

具体实施方式

[0033] 以下描述中,为了说明而不是为了限定,提出了诸如特定系统结构、技术之类的具体细节,以便透彻理解本申请实施例。然而,本领域的技术人员应当清楚,在没有这些具体细节的其它实施例中也可以实现本申请。在其它情况中,省略对众所周知的系统、装置、电路以及方法的详细说明,以免不必要的细节妨碍本申请的描述。

[0034] 应当理解,当在本申请说明书和所附权利要求书中使用时,术语“包括”指示所描述特征、整体、步骤、操作、元素和/或组件的存在,但并不排除一个或多个其它特征、整体、步骤、操作、元素、组件和/或其集合的存在或添加。

[0035] 还应当理解,在本申请说明书和所附权利要求书中使用的术语“和/或”是指相关联列出的项中的一个或多个的任何组合以及所有可能组合,并且包括这些组合。

[0036] 如在本申请说明书和所附权利要求书中所使用的那样,术语“如果”可以依据上下文被解释为“当...时”或“一旦”或“响应于确定”或“响应于检测到”。类似地,短语“如果确定”或“如果检测到[所描述条件或事件]”可以依据上下文被解释为意指“一旦确定”或“响应于确定”或“一旦检测到[所描述条件或事件]”或“响应于检测到[所描述条件或事件]”。

[0037] 另外,在本申请说明书和所附权利要求书的描述中,术语“第一”、“第二”、“第三”等仅用于区分描述,而不能理解为指示或暗示相对重要性。

[0038] 在本申请说明书中描述的参考“一个实施例”或“一些实施例”等意味着在本申请的一个或多个实施例中包括结合该实施例描述的特定特征、结构或特点。由此,在本说明书中的不同之处出现的语句“在一个实施例中”、“在一些实施例中”、“在其他一些实施例中”、“在另外一些实施例中”等不是必然都参考相同的实施例,而是意味着“一个或多个但不是所有的实施例”,除非是以其他方式另外特别强调。术语“包括”、“包含”、“具有”及它们的变

形都意味着“包括但不限于”,除非是以其他方式另外特别强调。

[0039] 在本申请实施例中,流程的执行主体包括终端设备。该终端设备包括但不限于:服务器、计算机、智能手机以及平板电脑等能够执行本申请公开的方法的设备。图1示出了本发明第一实施例公开的用于司法批量立案的智能化处理方法的流程示意图,详述如下:

[0040] S101,对批量案件材料进行多模态解析,提取出原始案件要素,通过跨模态信息互补机制对原始案件要素进行交叉验证,生成第一案件要素。

[0041] 在本实施例中,在法律实务中,面对大量的案件材料,传统的人工审查方式效率低下且容易出错。现有的一些案件要素提取技术大多仅基于单一模态的数据,如仅对文本进行关键词提取或仅对图像进行特征识别,难以全面、准确地获取案件的关键信息。不同模态的数据(如文本、图像、音频等)从不同角度反映了案件的信息,单一模态的处理方式忽略了各模态之间的关联和互补性,导致提取的案件要素可能存在遗漏或不准确的问题。因此,需要一种能够充分利用多模态数据信息,通过跨模态信息互补机制对原始案件要素进行交叉验证的方法,以提高案件要素提取的准确性和可靠性。

[0042] 具体的,采用自然语言处理技术,如词嵌入(WordEmbedding)、循环神经网络(RNN)及其变体(如长短期记忆网络LSTM、门控循环单元GRU)等,对案件文本进行分词、词性标注、命名实体识别等处理,提取出案件的关键信息,如当事人信息、案件事实、诉讼请求等。利用卷积神经网络(CNN)等深度学习模型,对案件相关的图像(如现场照片、证据图片等)进行特征提取,识别出图像中的关键物体、场景等信息。通过语音识别技术将案件相关的音频文件转换为文本,然后采用与文本解析类似的方法提取音频中的案件要素。

[0043] 分析不同模态数据之间的关联关系,例如,文本中提到的现场照片编号与图像文件之间的对应关系,音频中描述的案件事实与文本记录的一致性等。对于提取的原始案件要素,在不同模态之间进行交叉验证。例如,验证文本中提到的当事人姓名是否与图像中的人物特征相符,音频中描述的案件经过是否与文本记录一致等。根据交叉验证结果,对原始案件要素进行修正和补充。对于验证通过的要素予以保留,对于存在矛盾或不一致的要素进行进一步的分析和判断,确定最终的案件要素。将修正和补充后的案件要素整合为第一案件要素。

[0044] 在本实施例中,通过多模态解析和跨模态信息互补机制,充分利用了不同模态数据的信息,避免了单一模态处理的局限性,能够更全面、准确地提取案件要素。跨模态信息互补机制对原始案件要素进行了交叉验证,能够有效发现和纠正要素中的错误和矛盾,提高了案件要素的可靠性。本发明的方法能够自动对批量案件材料进行处理,大大减少了人工审查的工作量,提高了法律实务工作的效率。

[0045] S102,根据第一案件要素构建案件相似度矩阵,基于案件相似度矩阵,采用改进BERT-Siamese网络计算案件间的事实关联性权重,同步通过法律程序冲突检测算法识别管辖竞合与举证冲突,生成动态分组方案。

[0046] 在本实施例中,在法律实务中,面对大量的案件,如何高效、准确地对案件进行分组处理是一个关键问题。传统的案件分组方法主要依赖人工经验和简单的规则匹配,存在主观性强、效率低下且准确性不足的问题。随着案件数量的不断增加和案件复杂性的提高,这种传统方法已难以满足实际需求。现有的自动化分组方法大多仅考虑案件的表面特征,忽略了案件之间的内在关联性和法律程序上的冲突情况,导致分组结果不合理,可能影响

案件的处理效率和公正性。因此,需要一种能够综合考虑案件相似度、事实关联性和法律程序冲突的动态分组方法。

[0047] 具体的,对每个案件的第一案件要素进行特征提取,将案件要素表示为向量形式。计算任意两个案件向量之间的相似度,如余弦相似度,构建案件相似度矩阵。

[0048] 设计改进的BERT-Siamese网络结构,该网络以两个案件的向量作为输入,输出两个案件之间的事实关联性得分。使用大量标注好的案件数据对改进的BERT-Siamese网络进行训练,优化网络参数,使其能够准确计算案件间的事实关联性权重。比如,BERT-Siamese网络结构由两个共享权重的BERT模型和一个全连接层组成,将两个案件的向量分别输入到两个BERT模型中,得到两个案件的语义表示,将两个语义表示输入到全连接层,输出两个案件之间的事实关联性得分。使用5000份标注好的合同纠纷案件数据对改进的BERT-Siamese网络进行训练,经过多轮迭代后,网络能够准确计算合同纠纷案件间的事实关联性权重。

[0049] 制定法律程序冲突检测规则,包括管辖竞合规则和举证冲突规则。对每个案件对,根据其第一案件要素和法律程序冲突检测规则,判断是否存在管辖竞合或举证冲突。比如,如果两个案件的合同签订地或履行地在同一地区,且合同金额超过一定数额,则可能存在管辖竞合。如果两个案件的违约责任条款存在冲突,且双方当事人提供的证据相互矛盾,则可能存在举证冲突。对每对合同纠纷案件,根据其第一案件要素和上述规则,判断是否存在管辖竞合或举证冲突。

[0050] 综合考虑案件间的事实关联性权重和法律程序冲突情况,采用聚类算法(如K-means算法)对案件进行分组。在分组过程中,优先将事实关联性高且不存在法律程序冲突的案件分在同一组。比如,采用K-means算法对100份合同纠纷案件进行分组,K值根据实际情况设定为10。在分组过程中,优先将事实关联性得分高且不存在法律程序冲突的案件分在同一组。最终生成10个动态分组方案,每个分组中的案件具有较高的事实关联性,且不存在管辖竞合和举证冲突。

[0051] 在本实施例中,通过构建案件相似度矩阵和计算事实关联性权重,能够充分考虑案件之间的内在关联性,使分组结果更加准确合理。法律程序冲突检测算法能够及时发现管辖竞合和举证冲突,避免在后续处理过程中出现法律程序上的问题,提高案件处理的效率和公正性。动态分组方案能够根据案件的实际情况进行灵活调整,将相似度高且无冲突的案件分在同一组,便于集中处理,提高法律实务工作的效率。

[0052] S103,根据动态分组方案匹配分类文书模板库,调用Transformer-XL增强生成模型批量生成法律文书。

[0053] 在本实施例中,在法律实务中,法律文书的撰写是一项重要且繁琐的工作。传统的法律文书撰写方式主要依赖人工完成,这不仅效率低下,而且容易出现错误和不一致的情况。随着案件数量的不断增加,人工撰写法律文书的方式已难以满足实际需求。目前,虽然有一些自动化法律文书生成技术,但大多存在以下问题:一是无法根据案件的具体情况进行灵活分组和分类,导致生成的文书缺乏针对性;二是使用的生成模型在处理长文本和捕捉长距离依赖关系方面能力不足,生成的文书质量不高。因此,需要一种能够根据动态分组方案匹配分类文书模板库,并利用先进生成模型批量生成高质量法律文书的方法。

[0054] 具体的,构建一个包含多种类型法律文书模板的分类文书模板库,每个模板都有相应的标签和适用条件。根据动态分组方案中的案件类型、案件特征等信息,为每个分组匹

配最合适的文书模板。比如,构建一个合同纠纷相关的分类文书模板库,包含起诉状、答辩状、判决书等多种类型的文书模板。根据每个分组的案件类型(如买卖合同纠纷、租赁合同纠纷等)和案件特征(如合同金额、违约责任等),为每个分组匹配最合适的文书模板。例如,对于买卖合同纠纷且涉及较大金额的分组,匹配详细的起诉状和判决书模板。

[0055] 设计Transformer-XL增强生成模型,该模型在Transformer-XL的基础上进行了改进,增加了对法律领域知识的融合和长文本处理能力。将分组中的案件信息(如案件事实、法律依据、当事人信息等)和选定的文书模板进行预处理,转换为模型可接受的输入格式。将预处理后的输入数据输入到Transformer-XL增强生成模型中,模型根据输入信息生成相应的法律文书。比如,该模型在训练过程中融入了大量的合同纠纷案例数据和法律专业知识,能够更好地理解合同纠纷案件的特点和法律要求。对每个分组中的案件信息进行提取和整理,包括合同签订时间、履行情况、违约事实等,并与选定的文书模板进行拼接。将拼接后的数据输入到Transformer-XL增强生成模型中,模型根据输入信息生成相应的法律文书。例如,对于起诉状模板,模型会生成包含当事人信息、诉讼请求、事实与理由等内容的起诉状。

[0056] 对每个分组依次进行上述操作,批量生成与分组案件相对应的法律文书。对生成的法律文书进行质量检查和优化,确保文书的准确性和规范性。比如,对10个分组依次进行上述操作,批量生成100份合同纠纷相关的法律文书。对生成的法律文书进行质量检查,检查文书的格式是否规范、内容是否准确、法律依据是否充分等。对于存在问题的文书,进行相应的修改和优化。

[0057] 在本实施例中,通过根据动态分组方案匹配分类文书模板库,能够快速为每个分组选择合适的模板,并利用Transformer-XL增强生成模型批量生成法律文书,大大提高了文书生成的效率。动态分组方案考虑了案件的具体情况,匹配的文书模板能够更好地适应不同分组案件的特点,使生成的法律文书更具针对性。Transformer-XL增强生成模型具有强大的长文本处理能力和对法律知识的理解能力,能够生成准确、规范、高质量的法律文书,减少了人工撰写过程中可能出现的错误和不一致情况。

[0058] S104,通过关联图谱对批量的法律文书进行全流程智能校验,检测材料完整性、逻辑一致性及法律合规性,根据关联图谱的校验结果调整分组策略或文书内容。

[0059] 在本实施例中,在法律实务中,法律文书的准确性和合规性至关重要。随着法律业务的不断拓展,需要处理的法律文书数量日益增多,传统的人工校验方式不仅效率低下,而且容易出现遗漏和错误。同时,法律文书之间往往存在复杂的关联关系,人工校验难以全面把握这些关系,导致无法有效检测材料完整性、逻辑一致性及法律合规性。因此,需要一种能够利用关联图谱对批量法律文书进行全流程智能校验,并根据校验结果进行动态调整的方法。

[0060] 具体的,对批量法律文书进行实体抽取,识别出文书中的关键实体,如当事人、案件事实、法律依据等。分析实体之间的关系,构建实体之间的关联关系网络,形成法律文书关联图谱。比如,使用自然语言处理技术对100份法律文书进行实体抽取,识别出合同当事人、合同条款、违约事实等关键实体。分析实体之间的关系,如合同当事人与合同条款之间的签订关系、违约事实与合同条款之间的违反关系等,构建合同纠纷案件法律文书的关联图谱。

[0061] 检查关联图谱中是否存在缺失的实体或关系,判断法律文书是否包含了所有必要的材料。比如检查关联图谱中是否包含了所有合同当事人的信息、合同条款的详细内容以及违约事实的证据等。若发现某份法律文书中缺少了违约金额的计算依据,判定该文书材料不完整。

[0062] 分析关联图谱中实体和关系的逻辑关系,检测是否存在逻辑矛盾或不合理之处。比如,分析关联图谱中合同条款与违约事实之间的逻辑关系,判断是否存在逻辑矛盾。若发现某份法律文书中既认定了甲方违约,又认定了乙方违约,但在合同条款中并没有明确双方同时违约的处理方式,判定该文书逻辑不一致。

[0063] 将关联图谱中的实体和关系与法律法规进行比对,判断法律文书是否符合法律规定。比如,将关联图谱中的合同条款与相关法律法规进行比对,判断是否符合法律规定。若发现某份法律文书中的违约金计算方式超过了法律规定的上限,判定该文书不符合法律合规性要求。

[0064] 根据校验结果,对分组策略进行调整。将材料不完整、逻辑不一致或不符合法律合规性要求的法律文书所在的分组进行重新划分,将问题较为集中的文书归为一组,以便进行针对性的处理。对存在问题的法律文书内容进行修改和完善。对于材料不完整的文书,补充缺失的信息;对于逻辑不一致的文书,调整逻辑关系;对于不符合法律合规性要求的文书,修改合同条款或相关表述。

[0065] 在本实施例中,通过关联图谱对批量法律文书进行全流程智能校验,能够快速、准确地检测出材料完整性、逻辑一致性和法律合规性问题,大大提高了校验效率。关联图谱能够全面反映法律文书中的实体和关系,有效避免了人工校验的遗漏和错误,提高了校验的准确性。根据校验结果调整分组策略或文书内容,能够针对性地解决法律文书中存在的问题,提高法律文书的质量和合规性。

[0066] 在一些实施例中,上述步骤S101中,所述对批量案件材料进行多模态解析,具体包括:

[0067] 采用预构建的法律NLP模型对批量案件材料中的起诉状、证据清单进行实体识别与关系抽取,输出文本特征向量集;

[0068] 通过改进OCR-ResNet模型提取批量案件材料中的表格、签章及物证链,生成图像特征矩阵;

[0069] 使用ASR-Transformer模型将批量案件材料中的庭审录音转换为文本,提取出包含争议焦点的语音时序特征。

[0070] 在该实施例中,在法律实务中,案件材料通常以多种形式存在,包括起诉状、证据清单等文本材料,表格、签章、物证链等图像材料,以及庭审录音等语音材料。传统的案件材料处理方式往往针对单一模态进行,难以全面、准确地提取案件的关键信息。例如,仅对文本材料进行分析可能忽略图像和语音中的重要细节,导致对案件的理解不完整。因此,需要一种能够对批量案件材料进行多模态解析的方法,以充分挖掘案件材料中的潜在信息。

[0071] 具体的,利用预构建的法律NLP模型,该模型经过大量法律文本数据的训练,具备对法律术语和法律关系的理解能力。对起诉状和证据清单进行实体识别,识别出案件当事人、法律事实、法律依据等关键实体,并进行关系抽取,分析实体之间的逻辑关系。最终输出文本特征向量集,该向量集能够准确表示文本中的关键信息。比如,将50份起诉状和证据清

单输入预构建的法律NLP模型。模型识别出案件中的甲方、乙方、合同签订时间、违约条款等关键实体,并抽取了甲方违约、乙方要求赔偿等关系。输出文本特征向量集,每个向量代表一份起诉状或证据清单的关键信息。对于一份起诉状,输出的文本特征向量可能包含[甲方名称,乙方名称,违约金额,违约时间,诉讼请求]等信息。

[0072] 使用改进OCR-ResNet模型,该模型结合光学字符识别(OCR)技术和深度学习中的ResNet网络结构。首先,利用OCR技术对表格、签章等图像中的文字信息进行识别和提取;然后,通过ResNet网络对物证链等复杂图像进行特征提取,生成图像特征矩阵。该矩阵能够全面反映图像中的关键信息,包括文字内容、图像结构和特征等。比如,使用改进OCR-ResNet模型对20张合同表格图片进行处理。OCR技术识别出表格中的合同金额、付款方式等文字信息,ResNet网络提取出表格的结构特征,生成图像特征矩阵。对于签章图片,模型识别出签章的形状、颜色、字体等特征,并生成相应的特征向量。对于物证链图片,模型提取出物证的外观特征、摆放顺序等信息,生成特征矩阵。

[0073] 使用ASR-Transformer模型,该模型结合了自动语音识别(ASR)技术和Transformer网络结构。将庭审录音转换为文本后,利用Transformer模型的强大序列建模能力,提取出包含争议焦点的语音时序特征。这些特征能够准确反映庭审过程中双方当事人的争议点和辩论重点。比如,将5段庭审录音输入ASR-Transformer模型,转换为文本。模型提取出文本中的争议焦点,如合同是否有效、违约责任的承担等,并生成语音时序特征。对于一段庭审录音,输出的语音时序特征可能包含[争议焦点出现时间,争议焦点内容,双方当事人观点]等信息。

[0074] 在该实施例中,通过多模态解析,能够从文本、图像、语音等多种模态的案件材料中全面提取关键信息,避免了单一模态解析的局限性,提高了信息提取的完整性和准确性。自动化的多模态解析方法大大提高了案件材料分析的效率,减少了人工处理的时间和成本。生成的文本特征向量集、图像特征矩阵和语音时序特征为后续的法律分析和处理提供了便捷的数据支持。提取出的关键信息能够为法律专业人员提供更全面、准确的决策依据,有助于提高法律案件的处理质量和效率。例如,在合同纠纷案件中,通过分析文本、图像和语音特征,能够更准确地判断合同的有效性、违约责任的承担等问题。

[0075] 在一些实施例中,上述步骤S101中,所述通过跨模态信息互补机制对原始案件要素进行交叉验证,生成第一案件要素,具体包括:

[0076] 将原始案件要素按照要素类型分为事实要素、法律要素和程序要素;

[0077] 为每个原始案件要素分别添加来源模态标识,所述来源模态标识用于标识其来源于文本、图像或语音;

[0078] 基于来源模态标识,计算每个原始案件要素中涉及的文本特征向量集、图像特征矩阵和语音时序特征之间的余弦相似度,通过可学习权重参数动态调节文本特征向量集、图像特征矩阵和语音时序特征之间的模态贡献比例;

[0079] 当所述余弦相似度低于预设相似度阈值时,触发融合修正操作,将文本特征向量集、图像特征矩阵和语音时序特征进行加权融合,生成跨模态要素表示;

[0080] 基于要素类型和跨模态要素表示,将原始案件要素转换为第一案件要素。

[0081] 在该实施例中,在法律案件处理过程中,案件要素的准确性和完整性对于案件的判决结果至关重要。然而,原始案件要素往往来源于不同的模态,如文本(起诉状、证据清单

等)、图像(合同、物证照片等)和语音(庭审录音等)。不同模态的案件要素可能存在信息不一致或矛盾的情况,单纯依赖单一模态的案件要素进行法律分析和判断可能会导致错误的结果。因此,需要一种能够整合不同模态案件要素信息,进行交叉验证和修正的方法,以提高案件要素的质量。

[0082] 具体的,根据法律案件的特点,将原始案件要素分为事实要素(如案件事实经过、关键时间节点等)、法律要素(如法律依据、法律条款适用等)和程序要素(如诉讼程序、证据规则等)。

[0083] 为每个原始案件要素添加来源模态标识,例如,对于从文本中提取的案件要素,标识为“文本”;对于从图像中提取的案件要素,标识为“图像”;对于从语音中提取的案件要素,标识为“语音”。

[0084] 基于来源模态标识,获取每个原始案件要素对应的文本特征向量集、图像特征矩阵和语音时序特征。计算这些特征之间的余弦相似度,以衡量不同模态信息的一致性。同时,通过可学习权重参数动态调节不同模态特征在交叉验证中的贡献比例,以适应不同案件要素的特点。

[0085] 当余弦相似度低于预设相似度阈值时,认为不同模态的案件要素信息存在不一致或矛盾,触发融合修正操作。将文本特征向量集、图像特征矩阵和语音时序特征进行加权融合,生成跨模态要素表示。加权融合的权重根据可学习权重参数确定,以综合考虑不同模态信息的重要性。

[0086] 基于要素类型和跨模态要素表示,将原始案件要素转换为第一案件要素。第一案件要素具有更高的准确性和可靠性,能够更好地反映案件的真实情况。

[0087] 示例性的,假设有一批合同纠纷案件的原始案件要素,包括以下内容:事实要素,甲方于[具体日期]与乙方签订合同,约定甲方应在[具体日期]前支付货款[具体金额]。法律要素,根据《合同法》第[具体条款],乙方有权要求甲方承担违约责任。程序要素,本案应适用普通程序进行审理。将上述事实要素、法律要素和程序要素分别进行分类,并为每个要素添加来源模态标识。例如,事实要素可能来源于文本(合同文本、起诉状等),法律要素可能来源于文本(法律法规数据库),程序要素可能来源于文本(法院程序规定文件)。假设从文本中提取的事实要素特征向量为[向量A],从图像(合同扫描件)中提取的事实要素特征向量为[向量B],从语音(庭审录音)中提取的事实要素特征向量为[向量C]。计算[向量A]、[向量B]和[向量C]之间的余弦相似度。根据预设的可学习权重参数,动态调节文本、图像和语音模态在事实要素交叉验证中的贡献比例。例如,如果文本模态的信息较为丰富和准确,可适当提高其权重。如果计算得到的余弦相似度低于预设相似度阈值,例如0.8,说明不同模态的事实要素信息存在不一致。触发融合修正操作,将[向量A]、[向量B]和[向量C]进行加权融合,生成跨模态要素表示[向量D]。基于要素类型(事实要素)和跨模态要素表示[向量D],将原始事实要素转换为第一事实要素。例如,第一事实要素可能表述为“经综合验证,甲方于[具体日期]与乙方签订合同,约定甲方应在[具体日期]前支付货款[具体金额],但存在部分证据显示甲方可能存在延迟支付的情况”。

[0088] 在该实施例中,通过跨模态信息互补机制,对原始案件要素进行交叉验证和修正,能够有效发现和纠正不同模态信息之间的不一致和矛盾,提高案件要素的准确性。融合不同模态的信息,生成跨模态要素表示,使案件要素具有更强的可靠性和说服力,能够更好地

支持法律分析和判断。自动化的跨模态信息互补机制减少了人工核对和修正案件要素的时间和成本,提高了法律案件的处理效率。本发明的方法能够适应不同类型和复杂程度的法律案件,为各种案件提供准确的案件要素支持,有助于提高法律判决的质量和公正性。

[0089] 在一些实施例中,上述步骤S102中,所述根据第一案件要素构建案件相似度矩阵,具体包括:

[0090] 根据文本特征向量集、图像特征矩阵和语音时序特征之间的余弦相似度及对应的模态贡献比例,计算获得跨模态验证置信度;

[0091] 基于跨模态验证置信度的均值,分别计算事实要素、法律要素及程序要素的动态权重,形成案件相似度矩阵。

[0092] 在该实施例中,在法律领域,随着案件数量的不断增加,如何高效地对案件进行管理和分析成为了一个重要的问题。案件相似度分析是解决这一问题的关键技术之一。传统的案件相似度分析方法往往只考虑单一模态的案件要素,如文本信息,而忽略了图像、语音等其他模态信息的重要性。不同模态的案件要素包含了案件的不同方面信息,综合利用这些信息可以提高案件相似度分析的准确性和可靠性。因此,需要一种能够根据多种模态案件要素构建案件相似度矩阵的方法。

[0093] 具体的,对于每个第一案件要素,根据其对应的文本特征向量集、图像特征矩阵和语音时序特征,计算它们之间的余弦相似度。同时,考虑不同模态在案件要素中的贡献比例,通过加权求和的方式计算获得跨模态验证置信度。跨模态验证置信度反映了该案件要素在不同模态信息下的一致性和可靠性。基于所有第一案件要素的跨模态验证置信度的均值,分别计算事实要素、法律要素及程序要素的动态权重。动态权重反映了不同类型要素在案件相似度分析中的重要性。然后,根据这些动态权重和案件要素之间的相似度计算结果,构建案件相似度矩阵。

[0094] 示例性的,假设有一批合同纠纷案件,每个案件都包含了第一案件要素,包括事实要素(如合同签订时间、金额、履行情况等)、法律要素(如适用的法律法规、违约责任条款等)和程序要素(如诉讼程序、管辖法院等)。对于每个案件的事实要素,从文本(合同文本、起诉状等)、图像(合同扫描件、现场照片等)和语音(庭审录音、证人证言等)中提取特征向量、特征矩阵和时序特征。计算文本特征向量集、图像特征矩阵和语音时序特征之间的余弦相似度。例如,文本特征向量集与图像特征矩阵的余弦相似度为0.85,文本特征向量集与语音时序特征的余弦相似度为0.80,图像特征矩阵与语音时序特征的余弦相似度为0.78。假设文本模态的贡献比例为0.4,图像模态的贡献比例为0.3,语音模态的贡献比例为0.3。根据加权求和公式,计算该事实要素的跨模态验证置信度为:0.85×0.4+0.80×0.3+0.78×0.3=0.814。同样地,计算每个案件的法律要素和程序要素的跨模态验证置信度。计算所有案件的事实要素、法律要素和程序要素的跨模态验证置信度的均值。假设事实要素的均值为0.8,法律要素的均值为0.75,程序要素的均值为0.7。根据这些均值计算动态权重,例如,采用归一化方法,将事实要素的动态权重计算为0.8/(0.8+0.75+0.7)≈0.34,法律要素的动态权重计算为0.75/(0.8+0.75+0.7)≈0.32,程序要素的动态权重计算为0.7/(0.8+0.75+0.7)≈0.34。计算案件之间的相似度,例如,对于两个案件A和B,分别计算它们的事实要素相似度(如文本相似度、图像相似度等)、法律要素相似度和程序要素相似度。然后,根据动态权重对这些相似度进行加权求和,得到案件A和B之间的综合相似度。重复上述步骤,计算

所有案件之间的相似度,构建案件相似度矩阵。

[0095] 在该实施例中,通过综合考虑文本、图像和语音等多种模态的案件要素,计算跨模态验证置信度,并根据置信度均值计算动态权重,能够更准确地衡量案件之间的相似程度,避免了单一模态信息可能存在的片面性。基于准确的案件相似度矩阵,可以更高效地对案件进行分类和检索。例如,在法律数据库中,可以根据案件相似度快速找到与当前案件相似的历史案例,为法律判决提供参考。不同类型案件的事实要素、法律要素和程序要素在案件相似度分析中具有不同的重要性。通过计算动态权重,能够根据案件的具体情况,合理分配不同要素的权重,支持更复杂的法律分析任务。

[0096] 在一些实施例中,上述步骤S102中,所述基于案件相似度矩阵,采用改进BERT-Siamese网络计算案件间的事实关联性权重,同步通过法律程序冲突检测算法识别管辖竞合与举证冲突,生成动态分组方案,具体包括:

[0097] 将案件相似度矩阵输入至改进BERT-Siamese网络,通过双塔结构对案件要素对进行编码,获得编码案件特征;

[0098] 对编码案件特征进行多维权重融合,生成融合特征表示,采用混合相似度度量方法计算融合特征表示中的案件间的事实关联性权重;

[0099] 基于案件相似度矩阵,通过比较案件管辖法院是否一致,并结合案件当事人集合的交集判断冲突程度,获得第一冲突检测结果;

[0100] 基于案件相似度矩阵,通过比较案件举证期限的时间重叠度,并结合证据链相似度综合评估冲突强度,获得第二冲突检测结果;

[0101] 根据事实关联性权重、第一冲突检测结果和第二冲突检测结果,采用改进谱聚类算法生成动态分组方案。

[0102] 在该实施例中,在法律实践中,随着案件数量的急剧增加,如何对大量案件进行合理分组成为了一个亟待解决的问题。传统的案件分组方法往往仅基于案件的简单特征进行划分,忽略了案件之间复杂的事实关联性和法律程序冲突。不同案件之间可能存在事实上的紧密联系,也可能在管辖法院和举证程序上存在冲突。因此,需要一种能够综合考虑案件事实关联性和法律程序冲突的动态分组方法,以提高法律案件处理的效率和质量。

[0103] 具体的,利用改进的BERT-Siamese网络,其双塔结构分别对案件相似度矩阵中的案件要素对进行编码。通过对案件要素的深度语义理解,获得更具代表性的编码案件特征。对编码案件特征进行多维权重融合,综合考虑不同特征维度的重要性。采用混合相似度度量方法,结合余弦相似度、欧氏距离等多种度量方式,计算融合特征表示中的案件间的事实关联性权重。基于案件相似度矩阵,比较案件管辖法院是否一致。同时,结合案件当事人集合的交集情况,判断管辖竞合的程度。例如,若两个案件的管辖法院相同且当事人集合交集较大,则可能存在管辖竞合。比较案件举证期限的时间重叠度,并结合证据链相似度综合评估冲突强度。若两个案件的举证期限存在较大重叠,且证据链相似度较高,则可能存在举证冲突。根据事实关联性权重、第一冲突检测结果和第二冲突检测结果,采用改进谱聚类算法对案件进行分组。改进谱聚类算法能够更好地处理复杂的案件关系,生成更合理的动态分组方案。

[0104] 示例性的,假设有一批商业合同纠纷案件,每个案件都包含了详细的案件要素信息。将案件相似度矩阵输入至改进BERT-Siamese网络。网络对每个案件要素对进行编码,例

如,对于合同签订时间、金额、履行情况等要素进行深度语义编码。经过编码后,获得每个案件要素对的编码案件特征。这些特征包含了案件要素的丰富语义信息,能够更好地反映案件之间的相似性。对编码案件特征进行多维权重融合。例如,根据合同金额、履行情况等要素的重要性,为不同特征维度分配不同的权重。采用混合相似度度量方法计算案件间的事实关联性权重。假设两个案件A和B,通过余弦相似度和欧氏距离的综合计算,得到它们之间的事实关联性权重为0.75。比较案件A和B的管辖法院,发现它们属于同一法院管辖。同时,案件A和B的当事人集合交集较大,包含多个共同当事人。因此,判断案件A和B存在管辖竞合。比较案件A和B的举证期限,发现它们的举证期限存在较大重叠,重叠时间达到30天。此外,两个案件的证据链相似度较高,都涉及合同履行过程中的相关证据。因此,判断案件A和B存在举证冲突。根据事实关联性权重0.75、管辖竞合检测结果和举证冲突检测结果,采用改进谱聚类算法对案件进行分组。改进谱聚类算法考虑了案件之间的复杂关系,将案件A和B以及其他具有相似事实关联性和冲突情况的案件分为同一组。同时,对其他案件也进行了合理的分组,生成了动态分组方案。

[0105] 在该实施例中,通过改进BERT-Siamese网络计算案件间的事实关联性权重,结合法律程序冲突检测算法,能够更准确地衡量案件之间的关系,避免传统分组方法的片面性。动态分组方案可以根据案件的实际情况进行灵活调整,合理分配法律资源,提高法律处理的效率和质量。对于具有复杂事实关联性和法律程序冲突的案件,本发明的方法能够提供更合理的分组方案,有助于法律专业人员更好地处理案件。

[0106] 在一些实施例中,上述步骤S103中,所述根据动态分组方案匹配分类文书模板库,调用Transformer-XL增强生成模型批量生成法律文书,具体包括:

[0107] 通过计算动态分组方案中各案件组的特征向量与分类文书模板库的每个模板的特征锚点之间的加权余弦相似度,确定法律文书模板,所述特征锚点包括案件类型编码、标的额分段编码和程序阶段编码;

[0108] 构建Transformer-XL增强生成模型,所述Transformer-XL增强生成模型包括案件组级记忆单元和双通道生成单元,所述案件组级记忆单元用于通过记忆矩阵缓存同组案件的共性特征,并将记忆矩阵与当前案件隐藏状态拼接形成增强特征表示,所述双通道生成单元用于输入案件组共性特征和个案专属特征,并通过联合概率分布生成法律文书内容;

[0109] 将法律文书模板输入至Transformer-XL增强生成模型中,批量生成法律文书。

[0110] 在该实施例中,在法律实务中,法律文书的生成是一项繁琐且重要的工作。传统的法律文书生成方式主要依赖人工撰写,不仅效率低下,而且容易出现错误和不一致的情况。随着人工智能技术的发展,一些自动化法律文书生成方法逐渐出现,但这些方法往往难以处理具有复杂特征和关联性的案件组,生成的文书质量也有待提高。因此,需要一种能够根据案件的动态分组情况,精准匹配文书模板,并高效生成高质量法律文书的方法。

[0111] 具体的,利用动态分组方案中各案件组的特征向量与分类文书模板库的特征锚点进行加权余弦相似度计算,根据相似度结果确定最合适的法律文书模板。利用案件组级记忆用于缓存同组案件的共性特征,形成记忆矩阵,并与当前案件隐藏状态拼接,增强特征表示能力。利用双通道生成单元接收案件组共性特征和个案专属特征,通过联合概率分布生成准确、全面的法律文书内容。将匹配好的法律文书模板输入到Transformer-XL增强生成模型中,利用模型强大的生成能力,批量生成高质量的法律文书。

[0112] 示例性的,假设有一批民事借贷纠纷案件,经过前期的案件相似度分析和动态分组,得到了包含多个案件组的动态分组方案。计算每个案件组的特征向量,特征向量包含了案件类型编码(民事借贷纠纷)、标的额分段编码(如小额、中额、大额)和程序阶段编码(起诉阶段)等信息。将案件组特征向量与分类文书模板库中的每个模板的特征锚点进行加权余弦相似度计算。例如,对于案件组A,经过计算,与模板T1的加权余弦相似度为0.85,与模板T2的相似度为0.72,因此确定模板T1为最适合案件组A的法律文书模板。案件组级记忆单元对于案件组A中的所有案件,提取共性特征,如借款金额范围、还款期限等,形成记忆矩阵。将记忆矩阵与当前正在处理的案件的隐藏状态拼接,得到增强特征表示。双通道生成单元将案件组共性特征(如借贷双方的常见纠纷点)和个案专属特征(如具体的借款金额、还款日期等)输入到双通道生成单元中。通过联合概率分布,生成起诉状的内容。例如,生成的起诉状中包含了准确的借款人信息、借款金额、还款要求等内容。将模板T1输入到Transformer-XL增强生成模型中,利用模型对案件组A中的所有案件进行批量处理,生成了格式规范、内容准确的起诉状。

[0113] 在该实施例中,通过动态分组方案匹配模板和Transformer-XL增强生成模型的批量生成能力,大大缩短了法律文书的生成时间,提高了工作效率。基于案件特征向量和模板特征锚点的匹配方式,能够选择更合适的模板,结合Transformer-XL增强生成模型的强大生成能力,生成的法律文书内容准确、格式规范。能够处理不同类型、不同阶段的法律案件,根据案件的实际情况生成个性化的法律文书。

[0114] 在一些实施例中,上述步骤S104中,所述通过关联图谱对批量的法律文书进行全流程智能校验,检测材料完整性、逻辑一致性及法律合规性,具体包括:

[0115] 设置关联图谱,所述关联图谱的顶点集包括案件要素、法律条文、历史判例和法律文书,所述动态关联图谱的边集包括案件要素相似度、任一案件要素与任一引用法条的关系、任一法律文书与任一历史判例之间的逻辑冲突关系;

[0116] 根据关联图谱批量检测每个法律文书的案件要素与法律文书的映射覆盖率,获得材料完整性校验结果;

[0117] 根据关联图谱批量检测每个法律文书与每个历史判例之间的逻辑冲突关系,获得逻辑一致性校验结果;

[0118] 根据关联图谱批量检测每个法律文书的文书内容与每个引用法条之间的匹配度,获得法律合规性校验结果。

[0119] 在该实施例中,在法律实践中,法律文书的校验是一项至关重要的工作,需要确保材料的完整性、逻辑的一致性以及法律的合规性。传统的法律文书校验主要依赖人工审核,这种方式不仅效率低下,而且容易受到审核人员主观因素的影响,导致校验结果不准确。随着法律文书数量的不断增加,传统校验方式已难以满足实际需求。因此,需要一种智能化、自动化的法律文书校验方法。

[0120] 具体的,将案件要素、法律条文、历史判例和法律文书作为顶点,根据它们之间的关系构建边集,形成完整的关联图谱。通过计算法律文书中案件要素与关联图谱中相关要素的映射覆盖率,判断法律文书是否包含了所有必要的案件要素。分析法律文书与历史判例之间的逻辑冲突关系,检查法律文书的逻辑推理是否合理,是否存在与历史判例相矛盾的情况。评估法律文书的文书内容与引用法条之间的匹配度,确保法律文书的法律依据正

确、适用准确。

[0121] 示例性的,假设有一批民事合同纠纷的法律文书需要进行校验,具体步骤如下:构建关联图谱,包括顶点集和边集,其中,顶点集:案件要素包括合同主体、合同标的、合同履行情况等;法律条文包括《中华人民共和国民法典》中关于合同的相关条款;历史判例选取了近年来具有代表性的民事合同纠纷案例;法律文书即为待校验的这批民事合同纠纷法律文书。边集:计算不同合同主体之间的相似度,确定案件要素相似度边;建立合同要素与相关法律条文的引用关系边;分析法律文书与历史判例之间的逻辑关联,标记逻辑冲突关系边。对于每份法律文书,检查其中是否包含了合同主体、合同标的、合同履行情况等关键案件要素。例如,某份法律文书中缺少合同履行情况的描述,通过关联图谱可以直观地发现这一缺失,并得出该法律文书材料不完整的校验结果。将每份法律文书的逻辑推理与历史判例进行对比。例如,某份法律文书中关于合同违约责任的认定与某个历史判例中的认定方式存在明显差异,且这种差异缺乏合理的解释,通过关联图谱可以检测到这种逻辑冲突,并得出该法律文书逻辑不一致的校验结果。检查每份法律文书引用的法律条文是否与案件事实相符。例如,某份法律文书在处理合同解除问题时引用了不相关的法律条文,通过关联图谱可以评估出该法律文书的法律合规性存在问题,并给出具体的修改建议。

[0122] 在该实施例中,基于关联图谱的智能校验方法能够全面、准确地检测法律文书的材料完整性、逻辑一致性和法律合规性,避免了人工校验可能出现的疏漏和错误。批量处理法律文书,大大缩短了校验时间,提高了工作效率,能够满足大规模法律文书校验的需求。关联图谱涵盖了案件要素、法律条文、历史判例和法律文书等多个方面的信息,能够从多个角度对法律文书进行校验,确保校验结果的全面性和可靠性。

[0123] 在一些实施例中,上述步骤S104中,所述根据关联图谱的校验结果调整分组策略或文书内容,具体包括:

[0124] 基于材料完整性校验结果、逻辑一致性校验结果和法律合规性校验结果,确定错误边集合和错误特征模式;

[0125] 根据错误边集合重构案件相似度矩阵,根据错误特征模式对法律文书的文书内容进行修正处理。

[0126] 在该实施例中,在法律实务中,对大量法律文书进行处理时,通常需要先进行分组以便批量处理,同时要确保文书内容的准确性和合规性。然而,传统的方法在分组时可能存在不合理性,导致处理效率低下,且对文书内容的校验和修正往往缺乏系统性和准确性。随着法律文书数量的不断增加和复杂性的提高,传统方法已难以满足实际需求。因此,需要一种能够根据关联图谱校验结果智能调整分组策略和文书内容的方法。

[0127] 具体的,对关联图谱的校验结果进行综合分析,找出导致材料不完整、逻辑不一致和法律不合规的边,形成错误边集合。同时,总结这些错误所呈现出的共同特征,形成错误特征模式。利用错误边集合所反映的信息,重新计算案件之间的相似度,构建新的相似度矩阵,以便更合理地调整分组策略。根据错误特征模式,对法律文书中存在问题的内容进行针对性的修正,确保文书内容的准确性和合规性。

[0128] 示例性的,假设有一批商业合同纠纷的法律文书需要进行处理,具体步骤如下:关联图谱校验,包括材料完整性校验、逻辑一致性校验和法律合规性校验,其中,材料完整性校验:发现部分法律文书缺少合同签订日期、合同金额等关键信息。逻辑一致性校验:部分

法律文书在合同违约责任的认定上,与历史判例存在逻辑冲突。法律合规性校验:部分法律文书引用的法律条文与案件事实不匹配。将缺少关键信息的边、存在逻辑冲突的边以及法律条文引用错误的边收集起来,形成错误边集合。总结出材料缺失的常见要素、逻辑冲突的主要类型以及法律条文引用的常见错误模式。根据错误边集合,重新评估案件之间的相似度,例如,对于缺少相同关键信息的案件,降低它们之间的相似度;对于存在相同逻辑冲突类型的案件,增加它们之间的相似度。基于新的相似度评估结果,构建新的案件相似度矩阵。利用新的相似度矩阵,采用聚类算法对法律文书进行重新分组,例如,将具有相似逻辑冲突类型的案件分为一组,以便进行集中处理。根据错误特征模式,对法律文书中存在的问题进行修正,例如,补充缺少的合同签订日期和合同金额信息;调整合同违约责任的认定逻辑,使其与历史判例保持一致;修正引用的法律条文,确保其与案件事实相符。

[0129] 在该实施例中,基于关联图谱校验结果重构案件相似度矩阵,能够更准确地反映案件之间的相似性,从而使分组更加合理,提高批量处理的效率。根据错误特征模式对法律文书内容进行修正处理,能够及时发现并纠正文书中的错误,确保文书内容的准确性和合规性。整个过程实现了自动化和智能化,减少了人工干预,提高了法律文书处理的效率和质量。

[0130] 参照图2,本发明一实施例提供了一种用于司法批量立案的智能化处理系统2,所述系统2具体包括:

[0131] 第一处理模块201,用于对批量案件材料进行多模态解析,提取出原始案件要素,通过跨模态信息互补机制对原始案件要素进行交叉验证,生成第一案件要素;

[0132] 第二处理模块202,用于根据第一案件要素构建案件相似度矩阵,基于案件相似度矩阵,采用改进BERT-Siamese网络计算案件间的事实关联性权重,同步通过法律程序冲突检测算法识别管辖竞合与举证冲突,生成动态分组方案;

[0133] 第三处理模块203,用于根据动态分组方案匹配分类文书模板库,调用Transformer-XL增强生成模型批量生成法律文书;

[0134] 第四处理模块204,用于通过关联图谱对批量的法律文书进行全流程智能校验,检测材料完整性、逻辑一致性及法律合规性,根据关联图谱的校验结果调整分组策略或文书内容。

[0135] 可以理解的是,如图1所示的用于司法批量立案的智能化处理方法实施例中的内容均适用于本用于司法批量立案的智能化处理系统实施例中,本用于司法批量立案的智能化处理系统实施例所具体实现的功能与如图1所示的用于司法批量立案的智能化处理方法实施例相同,并且达到的有益效果与如图1所示的用于司法批量立案的智能化处理方法实施例所达到的有益效果也相同。

[0136] 需要说明的是,上述系统之间的信息交互、执行过程等内容,由于与本发明方法实施例基于同一构思,其具体功能及带来的技术效果,具体可参见方法实施例部分,此处不再赘述。

[0137] 所属领域的技术人员可以清楚地了解到,为了描述的方便和简洁,仅以上述各功能单元、模块的划分进行举例说明,实际应用中,可以根据需要而将上述功能分配由不同的功能单元、模块完成,即将所述系统的内部结构划分成不同的功能单元或模块,以完成以上描述的全部或者部分功能。实施例中的各功能单元、模块可以集成在一个处理单元中,也可

以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中,上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。另外,各功能单元、模块的具体名称也只是为了便于相互区分,并不用于限制本申请的保护范围。上述系统中单元、模块的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0138] 参照图3,本发明实施例还提供了一种计算机设备3,包括:存储器302和处理器301及存储在存储器302上的计算机程序303,当所述计算机程序303在处理器301上被执行时,实现如上述方法中任一项所述的用于司法批量立案的智能化处理方法。

[0139] 所述计算机设备3可以是桌上型计算机、笔记本、掌上电脑及云端服务器等计算设备。该计算机设备3可包括,但不仅限于,处理器301、存储器302。本领域技术人员可以理解,图3仅仅是计算机设备3的举例,并不构成对计算机设备3的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件,例如还可以包括输入输出设备、网络接入设备等。

[0140] 所称处理器301可以是中央处理单元(Central Processing Unit,CPU),该处理器301还可以是其他通用处理器、数字信号处理器(Digital Signal Processor,DSP)、专用集成电路(Application Specific Integrated Circuit,ASIC)、现成可编程门阵列(Field-Programmable Gate Array,FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

[0141] 所述存储器302在一些实施例中可以是所述计算机设备3的内部存储单元,例如计算机设备3的硬盘或内存。所述存储器302在另一些实施例中也可以是所述计算机设备3的外部存储设备,例如所述计算机设备3上配备的插接式硬盘,智能存储卡(Smart Media Card,SMC),安全数字(Secure Digital,SD)卡,闪存卡(Flash Card)等。进一步地,所述存储器302还可以既包括所述计算机设备3的内部存储单元也包括外部存储设备。所述存储器302用于存储操作系统、应用程序、引导装载程序(Boot Loader)、数据以及其他程序等,例如所述计算机程序的程序代码等。所述存储器302还可以用于暂时地存储已经输出或者将要输出的数据。

[0142] 本发明实施例还提供了一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器运行时,实现如上述方法中任一项所述的用于司法批量立案的智能化处理方法。

[0143] 该实施例中,所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用时,可以存储在一个计算机可读取存储介质中。基于这样的理解,本申请实现上述实施例方法中的全部或部分流程,可以通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一计算机可读存储介质中,该计算机程序在被处理器执行时,可实现上述各个方法实施例的步骤。其中,所述计算机程序包括计算机程序代码,所述计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质至少可以包括:能够将计算机程序代码携带到拍照装置/终端设备的任何实体或装置、记录介质、计算机存储器、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、电载波信号、电信信号以及软件分发介质。例如U盘、移动硬盘、磁碟或者光盘等。在某些司法管辖区,根据立法和专利实践,计算机可读介质不可以是电载波

信号和电信信号。

[0144] 在上述实施例中,对各个实施例的描述都各有侧重,某个实施例中没有详述或记载的部分,可以参见其它实施例的相关描述。

[0145] 本领域普通技术人员可以意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本申请的范围。

[0146] 在本申请所公开的实施例中,应该理解到,所揭露的装置/终端设备和方法,可以通过其它的方式实现。例如,以上所描述的装置/终端设备实施例仅仅是示意性的,例如,所述模块或单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通讯连接可以是通过一些接口,装置或单元的间接耦合或通讯连接,可以是电性,机械或其它的形式。

[0147] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

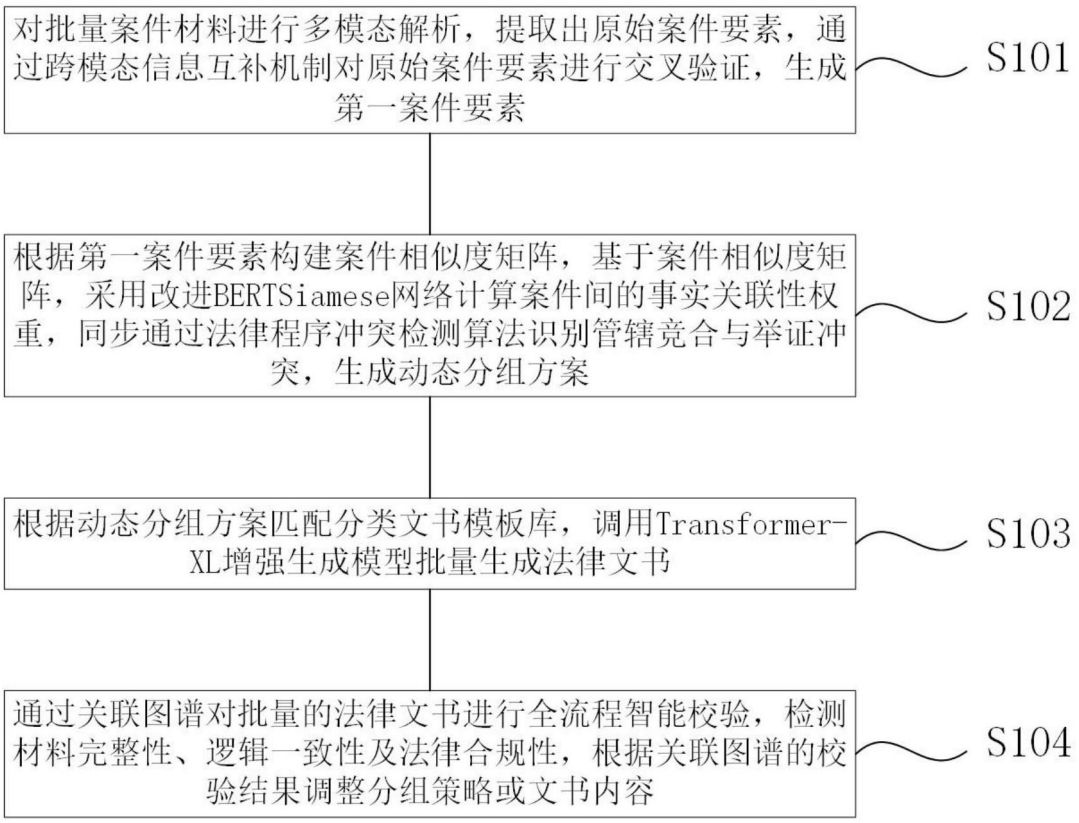

对批量案件材料进行多模态解析，提取出原始案件要素，通过跨模态信息互补机制对原始案件要素进行交叉验证，生成第一案件要素
S101
根据第一案件要素构建案件相似度矩阵，基于案件相似度矩阵，采用改进BERTSiamese网络计算案件间的事实关联性权重，同步通过法律程序冲突检测算法识别管辖竞合与举证冲突，生成动态分组方案
S102
根据动态分组方案匹配分类文书模板库，调用Transformer-XL增强生成模型批量生成法律文书
S103
通过关联图谱对批量的法律文书进行全流程智能校验，检测材料完整性、逻辑一致性及法律合规性，根据关联图谱的校验结果调整分组策略或文书内容
S104

图1

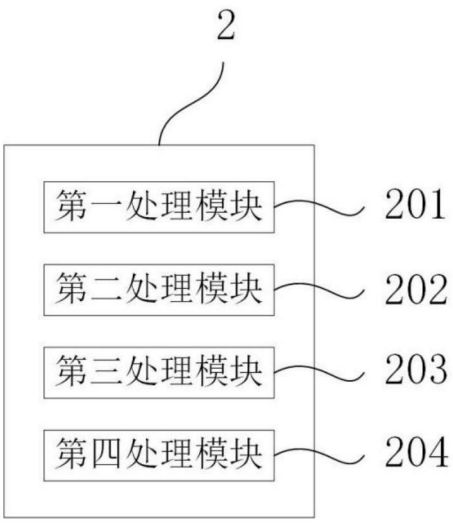

2
第一处理模块
201
第二处理模块
202
第三处理模块
203
第四处理模块
204

图2

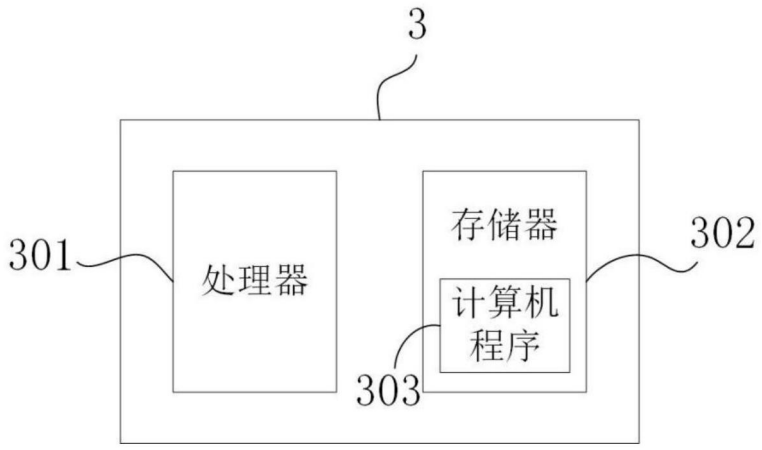


图3