

基于知识图谱的案件检索方法、装置、设备及存储介质

本申请要求于2020年1月8日提交中国专利局、申请号为202010017590.4，发明名称为“基于知识图谱的案件检索方法、装置、设备及存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

5 技术领域

本申请涉及大数据领域中的知识图谱领域，尤其涉及基于知识图谱的案件检索方法、装置、设备及存储介质。

背景技术

类似案件搜索方式是根据案件的相关信息对过往案件进行检索，找到最相似的案件的一种案件信息获取方式，所获得的案件信息能够为法官在判决时或诉讼案件的相关人员提供参考。现采用的类似案件搜索方式是仅基于检索裁判文书中的相关字段对裁判文书进行全文检索，没有处理裁判文书中没有司法裁判的相关知识的明确的对应关系、司法裁判的“关键事实”的复杂逻辑难以用关键字描述以及争议焦点、案由和诉请等相关信息在不同文书中的表述方式并不一致的影响因素，以致检索的准确性和专业性达不到较好的效果。

目前的类案检索中，通过接收用户输入的查询内容，对所述查询内容进行分析以得到所述查询内容的第一关键信息，根据所述第一关键信息从已创建的裁判知识图谱中获取第二关键信息，通过所述第二关键信息在数据库中检索，获得与所述第二关键信息关联的案例文本，对所述案例文本进行排序，并将经过排序的案例文本按照预设推荐条件进行推荐输出。发明人意识到，由于裁判知识图谱仅是根据裁判文书的内容进行构建的知识图谱，导致搜索的参考对象比较单一，导致所获取的案例文本作为参考性不大，以使检索次数和检索方式增加，因而，导致案件检索系统的可用性低。

技术问题

本申请提供了一种基于知识图谱的案件检索方法、装置、设备及存储介质，用于构建法律案件知识图谱以及对法律案件知识图谱的检索、随机游走采样和余弦相似度计算等步骤，使其法律案件知识图谱和待检索知识图谱的案件信息在描述上更具联系性、专业性和全面性，以及使系统具备高可拓展性和高操作性，以提高检索的速度和精准性，减少用户对系统的操作次数提高案件检索系统的可用性。

技术方案

本申请实施例的第一方面提供一种基于知识图谱的案件检索方法，包括：
通过预置的模型分析文本信息构建法律案件知识图谱，并分析所述法律案件知识图谱构建节点集合数据，所述文本信息包括历史和/或实时的原告诉请信息、被告的辩称信息、原告提供的证据信息、被告提供的证据信息和历史法律案件信息；

将所述节点集合数据的各节点作为出发点，对所述节点集合数据进行随机游走采样，获得多个序列数据；

基于多个所述序列数据通过词转化向量算法对所述模型进行训练，获取更新的目标模型；

获取目标文本信息，并通过所述目标模型分析所述目标文本信息，构建待检索知识图谱；

在所述法律案件知识图谱中检索，获取与所述待检索知识图谱关联的案件信息和第一词嵌入向量数据，以及获取所述待检索知识图谱的第二词嵌入向量数据；

根据所述第一词嵌入向量数据和所述第二词嵌入向量数据计算所述案件信息的第一相似度和第二相似度，并将所述第一相似度和所述第二相似度相加，获得目标相似度，所述第一相似度指示所述案件信息在内容描述上的相似度，所述第二相似度指示所述案件信息在逻辑关系上的相似度；

按照所述目标相似度的值从高到低对所述案件信息进行排序，并输出经过排序的案件信息。

本申请实施例的第二方面提供一种基于知识图谱的案件检索装置，具有实现对应于上述第一方面提供的基于知识图谱的案件检索方法的功能。所述功能可以通过硬件实现，也可以通过硬件执行相应的软件实现。硬件或软件包括一个或多个与上述功能相对应的模块，所述单元可以是软件和/或硬件。

所述装置包括：

第一构建单元，用于通过预置的模型分析文本信息构建法律案件知识图谱，并分析所述法律案件知识图谱构建节点集合数据，所述文本信息包括历史和/或实时的原告诉请信息、被告的辩称信息、原告提供的证据信息、被告提供的证据信息和历史法律案件信息；

采样单元，用于将所述节点集合数据的各节点作为出发点，对所述节点集合数据进行随机游走采样，获得多个序列数据；

第一获取单元，用于基于多个所述序列数据通过词转化向量算法对所述模型进行训练，获取更新的目标模型；

第二构建单元，用于获取目标文本信息，并通过所述目标模型分析所述目标文本信息，构建待检索知识图谱；

第二获取单元，用于在所述法律案件知识图谱中检索，获取与所述待检索知识图谱关联的案件信息和第一词嵌入向量数据，以及获取所述待检索知识图谱的第二词嵌入向量数据；

计算单元，用于根据所述第一词嵌入向量数据和所述第二词嵌入向量数据计算所述案件信息的第一相似度和第二相似度，并将所述第一相似度和所述第二相似度相加，获得目标相似度，所述第一相似度指示所述案件信息在内容描述上的相似度，所述第二相似度指示所述案件信息在逻辑关系上的相似度；

排序单元，用于按照所述目标相似度的值从高到低对所述案件信息进行排序，并输出经过排序的案件信息。

本申请实施例的第三方面提供了一种基于知识图谱的案件检索设备，包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现一种基于知识图谱的案件检索方法，包括：

通过预置的模型分析文本信息构建法律案件知识图谱，并分析所述法律案件知识图谱构建节点集合数据，所述文本信息包括历史和/或实时的原告诉请信息、被告的辩称信息、原告提供的证据信息、被告提供的证据信息和历史法律案件信息；

将所述节点集合数据的各节点作为出发点，对所述节点集合数据进行随机游走采样，获得多个序列数据；

基于多个所述序列数据通过词转化向量算法对所述模型进行训练，获取更新的目标模型；

获取目标文本信息，并通过所述目标模型分析所述目标文本信息，构建待检索知识图谱；

在所述法律案件知识图谱中检索，获取与所述待检索知识图谱关联的案件信息和第一词嵌入向量数据，以及获取所述待检索知识图谱的第二词嵌入向量数据；

根据所述第一词嵌入向量数据和所述第二词嵌入向量数据计算所述案件信息的第一相似度和第二相似度，并将所述第一相似度和所述第二相似度相加，获得目标相似度，所述第一相似度指示所述案件信息在内容描述上的相似度，所述第二相似度指示所述案件信息在逻辑关系上的相似度；

按照所述目标相似度的值从高到低对所述案件信息进行排序，并输出经过排序的案件信息。

本申请实施例的第四方面提供了一种计算机可读存储介质，包括指令，当所述指令在

计算机上运行时，使得计算机执行一种基于知识图谱的案件检索方法，包括：

通过预置的模型分析文本信息构建法律案件知识图谱，并分析所述法律案件知识图谱构建节点集合数据，所述文本信息包括历史和/或实时的原告诉请信息、被告的辩称信息、原告提供的证据信息、被告提供的证据信息和历史法律案件信息；

- 5 将所述节点集合数据的各节点作为出发点，对所述节点集合数据进行随机游走采样，获得多个序列数据；

基于多个所述序列数据通过词转化向量算法对所述模型进行训练，获取更新的目标模型；

- 10 获取目标文本信息，并通过所述目标模型分析所述目标文本信息，构建待检索知识图谱；

在所述法律案件知识图谱中检索，获取与所述待检索知识图谱关联的案件信息和第一词嵌入向量数据，以及获取所述待检索知识图谱的第二词嵌入向量数据；

- 15 根据所述第一词嵌入向量数据和所述第二词嵌入向量数据计算所述案件信息的第一相似度和第二相似度，并将所述第一相似度和所述第二相似度相加，获得目标相似度，所述目标相似度指示所述案件信息在内容描述上的相似度，所述第二相似度指示所述案件信息在逻辑关系上的相似度；

按照所述目标相似度的值从高到低对所述案件信息进行排序，并输出经过排序的案件信息。

有益效果

- 20 相较于现有技术，本申请实施例提供的技术方案中，通过根据文本信息构建法律案件知识图谱，将根据所述法律案件知识图谱构建的节点集合数据进行随机游走采样，获取多个序列数据，基于多个所述序列数据通过词转化向量算法对所述模型进行训练，获取更新的目标模型，获取目标文本信息，并通过所述目标模型分析所述目标文本信息以构建待检索知识图谱，根据所述待检索知识图谱在所述法律案件知识图谱中检索，以获取与所述待检索知识图谱关联的案件信息，根据所述案件信息的第一相似度和第二相似度，获取输出的案件信息。本申请实施例，使其法律案件知识图谱和待检索知识图谱的案件信息在描述上更具联系性、专业性和全面性，以及使系统具备高可拓展性和高操作性，以提高检索的速度和精准性，减少用户对系统的操作次数提高案件检索系统的可用性。

附图说明

- 30 图1为本申请实施例中基于知识图谱的案件检索方法的一个实施例示意图；
图2为本申请实施例中构建的法律案件知识图谱的一种示意图；
图3为本申请实施例中获取目标相似度中对比的待检索知识图谱的一种示意图；
图4为本申请实施例中获取目标相似度中对比的法律案件知识图谱的一种示意图；
图5为本申请实施例中基于知识图谱的案件检索方法的另一个实施例示意图；
35 图6为本申请实施例中基于知识图谱的案件检索装置的一个实施例示意图；
图7为本申请实施例中基于知识图谱的案件检索装置的另一个实施例示意图；
图8为本申请实施例中基于知识图谱的案件检索设备的一个实施例示意图。

本申请的最佳实施方式

- 请参阅图 1，本申请实施例提供的一种基于知识图谱的案件检索方法的流程图，以下
40 对本申请提供一种基于知识图谱的案件检索方法进行举例说明，该方法由计算机设备执行，计算机设备可为服务器或者终端，当图 8 所示的装置 80 为应用或者执行程序时，终端为安装图 8 所示的装置 80 的终端，本申请不对执行主体的类型作限制，具体包括：

- 101、通过预置的模型分析文本信息构建法律案件知识图谱，并分析法律案件知识图谱构建节点集合数据，文本信息包括历史和/或实时的原告诉请信息、被告的辩称信息、原告提供的证据信息、被告提供的证据信息和历史法律案件信息。
45

服务器通过预置的模型分析文本信息以构建法律案件知识图谱，并分析法律案件知识图谱以构建节点集合数据，文本信息包括历史和/或实时的原告诉请信息、被告的辩称信息、原告提供的证据信息、被告提供的证据信息和历史法律案件信息。

具体的，服务器对文本信息进行以清除文本信息中的噪声，纠正不一致，对文本信息进行数据集成，将文本信息有多个数据源合并成一个一致的数据存储，对文本信息进行聚集、删除冗余特征和聚类处理，对文本信息进行数据变换处理，以将文本信息压缩到较小的区间和机器能识别的表达式，以及对文本信息进行分词处理，对分词进行特征提取，获取特征信息，调用预置的卷积神经网络语言模型通过自然语言处理（Natural Language Processing, NLP）算法将特征信息解释为知识图谱的结构化数据，结构化数据包括属性、属性对应的内容以及属性之间的关系，根据属性、属性对应的内容以及属性之间的关系创建法律案件知识图谱，如图 2 所示。分析法律案件知识图谱，按照原告信息、原告的诉请信息、争论焦点、事实要素、小要素、发条和原告提供的证据信息，以及之前的关联关系，对法律案件知识图谱进行分拆和再组合，以将法律案件知识图谱的节点数据构建成树形结构的节点集合数据。图 2 内容仅作举例参考，其内容准确性和实际操作与否不作考虑。

102、将节点集合数据的各节点作为出发点，对节点集合数据进行随机游走采样，获得多个序列数据。

服务器将节点集合数据的各节点作为出发点，对节点集合数据进行随机游走采样，获得多个序列数据。

具体的，将法律案件知识图谱深度输入游走 Deepwalk 算法模型中对进行无监督的预训练。设置迭代次数 n 、步长 m 、变量数目 p 和每次随机生成向量的数量 q ，初始化随机游走次数 k ，以节点集合数据的各节点作为出发点，以出发点的下游节点集合的权重值对节点集合数据随机游走 m 步，进行随机游走 k 次，当随机游走 m 步时，当前所处节点没有下游连接点，则在该节点停止随机游走，并根据变量数据 p 和每次随机生成向量的数量 q 生成一条条的序列数据，迭代操作 n 次，当一系列的随机游走操作完成后，计算序列数据的更优值，若是无法获取到更优值，则重新设置步长，对节点集合数据进行随机游走采样，直至获取到更优值。其中，可通过计算目标函数的相反数的最小值来计算更优值，测试函数

$$\text{如下: } f = \frac{\sin\left(\sqrt{(x-\tau)^2 + (y-\tau)^2} + e\right)}{\sqrt{(x-\tau)^2 + (y-\tau)^2} + e} + 1, 0 \leq x, y \leq 2\tau, x \text{ 与 } y \text{ 表示随机游走停止时所处的节点的坐标, } \tau \text{ 表示全局最大值取值范围, } e \text{ 表示控制精度。}$$

103、基于多个序列数据通过词转化向量算法对模型进行训练，获取更新的目标模型。服务器基于多个序列数据通过词转化向量算法对模型进行训练，获取更新的目标模型。

具体的，通过词转化向量 word2vec 算法对多个序列数据进行分析，以对模型进行训练，更新模型。可通过输入层将多个序列数据中的词转换成向量的表示形式，得到词向量，隐层计算输入层输出的词向量的累加和，得到一个总的向量，总的向量与输出层的霍夫曼树的每一个非叶节点相连接，根据语义关系和上下文关系对总的向量构建霍夫曼树，通过能量函数对霍夫曼树进行概率计算。

104、获取目标文本信息，并通过目标模型分析目标文本信息，构建待检索知识图谱。

服务器获取目标文本信息，并通过目标模型分析目标文本信息以构建待检索知识图谱。

具体的，服务器对目标文本信息进行以清除文本信息中的噪声，纠正不一致，对目标文本信息进行数据集成，将目标文本信息有多个数据源合并成一个一致的数据存储，对目标文本信息进行聚集、删除冗余特征和聚类处理，对目标文本信息进行数据变换处理，以将目标文本信息压缩到较小的区间和机器能识别的表达式，以及对目标文本信息进行分词处理，对分词进行特征提取，获取特征信息，调用预置的卷积神经网络语言模型通过 NLP 算法将特征信息解释为知识图谱的结构化数据，结构化数据包括属性、属性对应的内容以

及属性之间的关系,根据属性、属性对应的内容以及属性之间的关系创建待检索知识图谱。

105、在法律案件知识图谱中检索,获取与待检索知识图谱关联的案件信息和第一词嵌入向量数据,以及获取待检索知识图谱的第二词嵌入向量数据。

5 服务器在法律案件知识图谱中检索,获取与待检索知识图谱关联的案件信息和第一词嵌入向量数据,以及获取待检索知识图谱的第二词嵌入向量数据。

具体的,对待检索知识图谱进行特征提取,以及分析特征信息之间的联系,得到关联特征信息,关联特征信息包括案由信息和案件类型信息,根据关联特征信息遍历法律案件知识图谱,获取相同和/或相似的案件信息和案件信息对应的第一词嵌入向量数据,以及待检索知识图谱的第二词嵌入向量数据。以便于通过第一词嵌入向量数据和第二词嵌入向量数据计算待检索知识图谱的节点与法律案件知识图谱的节点之前的关系的相似度和内容的相似度。

106、根据第一词嵌入向量数据和第二词嵌入向量数据计算案件信息的第一相似度和第二相似度,并将第一相似度和第二相似度相加,获得目标相似度,第一相似度指示案件信息在内容描述上的相似度,第二相似度指示案件信息在逻辑关系上的相似度。

15 服务器根据第一词嵌入向量数据和第二词嵌入向量数据计算案件信息的在内容描述上的第一相似度和在逻辑关系上的第二相似度,并将第一相似度和第二相似度相加,获得目标相似度。

具体的,计算第一词嵌入向量数据和第二词嵌入向量数据之间在内容描述方面上的相似度,得到案件信息的第一相似度,计算第一词嵌入向量数据和第二词嵌入向量数据之间在逻辑关系方面上的相似度,得到案件信息的第二相似度,将第一相似度和第二相似度的和值作为目标相似度。

20 通过将法律案件知识图谱中多个节点对应的相似度相加所得值作为法律案件知识图谱对应的法律案件与用户输入的待检索的案件的相似度,以提高检索的准确性。例如:如图3和图4所示,图3为待检索知识图谱,图4为法律案件知识图谱,若图4中将节点6、节点8、节点9和节点10分别对应的第二相似度的和值与图3中的节点6、节点8、节点9和节点10分别对应的相似度的和值最为接近,则图4中的节点7就会与图3中的节点7最为相似。

上述举例内容、图3内容和图4内容仅作参考,其内容准确性和实际操作与否不作考虑。

30 107、按照目标相似度的值从高到低对案件信息进行排序,并输出经过排序的案件信息。

服务器按照目标相似度的值从高到低对案件信息进行排序,并输出经过排序的案件信息。

具体的,按照目标相似度的值从高到低对案件信息进行排序,得到排序的案件信息,对经过排序的案件进行解释说明,解释说明的内容包括案件信息与输入的文本信息的相似度、案件信息的判决意见和案件信息的证据提取方向等,并将案件信息和解释说明生成文本信息,以及对案件信息中各部分的相似度进行统计分析,生成可视化图表。当接收到用户输入的案件信息输出条件时,输出满足案件信息输出条件的案件信息的文本和可视化图表,案件输出条件包括但不限于案件信息的排序范围或案件的相似度范围。

35 请参阅图5,本申请实施例中基于知识图谱的案件检索方法的另一个实施例包括:

40 501、通过预置的模型分析文本信息构建法律案件知识图谱,并分析法律案件知识图谱构建节点集合数据,文本信息包括历史和/或实时的原告告诉请信息、被告的辩称信息、原告提供的证据信息、被告提供的证据信息和历史法律案件信息。

服务器通过预置的模型分析文本信息以构建法律案件知识图谱,并分析法律案件知识图谱以构建节点集合数据,文本信息包括历史和/或实时的原告告诉请信息、被告的辩称信息、原告提供的证据信息、被告提供的证据信息和历史法律案件信息。

45 具体的,服务器对文本信息进行以清除文本信息中的噪声,纠正不一致,对文本信息

进行数据集成，将文本信息有多个数据源合并成一个一致的数据存储，对文本信息进行聚集、删除冗余特征和聚类处理，对文本信息进行数据变换处理，以将文本信息压缩到较小的区间和机器能识别的表达式，以及对文本信息进行分词处理，对分词进行特征提取，获取特征信息，调用预置的卷积神经网络语言模型通过自然语言处理（Natural Language Processing, NLP）算法将特征信息解释为知识图谱的结构化数据，结构化数据包括属性、属性对应的内容以及属性之间的关系，根据属性、属性对应的内容以及属性之间的关系创建法律案件知识图谱，如图 2 所示。分析法律案件知识图谱，按照原告信息、原告的诉请信息、争论焦点、事实要素、小要素、发条和原告提供的证据信息，以及之前的关联关系，对法律案件知识图谱进行分拆和再组合，以将法律案件知识图谱的节点数据构建成树形结构的节点集合数据。图 2 内容仅作参考，其内容准确性和实际操作与否不作考虑。

可选的，通过预置的模型分析文本信息以构建法律案件知识图谱，具体包括：创建区块链结构网络，获取区块链结构网络存储的文本信息，区块链用于链接各个法律案件查询和/或管理平台的数据；对文本信息进行词法分析、句法分析以及语义分析，得到结构化数据，并将结构化数据转换为标记语言数据；对标记语言数据进行实体抽取以获取实体信息，以及对标记语言数据进行关系抽取得到关系信息，以及对标记语言数据进行属性抽取得到属性信息，实体信息包括法律案件涉及的人名、组织/机构名、地理位置、事件/日期、字符值和金额值，关系信息包括事实要素、人物、证据、争议焦点和法条中至少两者之间的关系；对实体信息、关系信息和属性信息进行信息融合处理，以获取融合处理信息，根据融合处理信息构建数据模型得到法律案件知识图谱，并将法律案件知识图谱存储至区块链结构网络的关联数据库管理系统 MySQL 数据库。通过区块链结构网络获取多个平台的历史数据和实时数据，为法律案件知识图谱提供稳健的构建基础，结合区块链结构网络的低成本、高效率和数据存储安全等的特性，使其法律案件知识图谱的构建快速而准确。

502、将节点集合数据的各节点作为出发点，对节点集合数据进行随机游走采样，获得多个序列数据。

服务器将节点集合数据的各节点作为出发点，对节点集合数据进行随机游走采样，获得多个序列数据。

具体的，将法律案件知识图谱深度输入游走 Deepwalk 算法模型中对进行无监督的预训练。设置迭代次数 n 、步长 m 、变量数目 p 和每次随机生成向量的数量 q ，初始化随机游走次数 k ，以节点集合数据的各节点作为出发点，以出发点的下游节点集合的权重值对节点集合数据随机游走 m 步，进行随机游走 k 次，当随机游走 m 步时，当前所处节点没有下游连接点，则在该节点停止随机游走，并根据变量数据 p 和每次随机生成向量的数量 q 生成一条条的序列数据，迭代操作 n 次，当一系列的随机游走操作完成后，计算一条条序列数据的更优值，若是无法获取到更优值，则重新设置步长，对节点集合数据进行随机游走采样，直至获取到更优值。其中，可通过计算目标函数的相反数的最小值来计算更优

值，测试函数如下：
$$f = \frac{\sin\left(\sqrt{(x-\tau)^2 + (y-\tau)^2 + e}\right)}{\sqrt{(x-\tau)^2 + (y-\tau)^2 + e}} + 1, 0 \leq x, y \leq 2\tau, x \text{ 与 } y \text{ 表示随机游走停止时所处的节点的坐标, } \tau \text{ 表示全局最大值取值范围, } e \text{ 表示控制精度。}$$

可选的，将节点集合数据的各节点作为出发点，对节点集合数据进行随机游走采样，获得多个序列数据，具体包括：将节点集合数据的各节点作为出发点，并获取出发点的下一节点的权重值；分析权重值以设置随机游走步数，以及分析出发点所在分支的节点数以设置游走阈值步数；当判断到随机游走步数小于游走阈值步数时，按照随机游走步数沿着连接出发点的下一节点的节点进行游走，并生成随机游走序列数据，获取多个序列数据；当判断到随机游走步数大于或等于游走阈值步数时，按照游走阈值步数沿着连接出发点的下一节点的节点进行游走，并生成随机游走序列数据，获取多个序列数据。

可选的, 将节点集合数据的各节点作为出发点, 对节点集合数据进行随机游走采样, 获得多个序列数据, 具体包括: 以节点集合数据的每个节点作为出发点, 获取除出发点之外的下游节点的权重值, 根据权重值确定当前迭代次数; 根据当前迭代次数, 按照预设路径长度对节点集合数据进行随机游走; 当检测到当前迭代次数小于预设迭代控制次数时, 5 随机生成区间为 $[-1, 1]$ 之前的多维向量以及多维向量所在的多元函数; 计算多元函数的函数值, 若检测到函数值达到第一预设阈值, 以及预设路径长度小于控制精度, 则停止对节点集合数据随机游走, 并获取在节点集合数据中进行随机游走时所生成的序列数据, 控制精度用于停止对节点集合数据随机游走; 若检测到函数值未达到第一预设阈值, 以及预设路径长度不小于控制精度, 则设置随机游走的行走步长为预设路径长度的一半, 按照设置的随机游走的行走步长重新对节点集合数据随机游走, 直至函数值达到第一预设阈值以及 10 预设路径长度小于控制精度, 并获取在节点集合数据中进行随机游走时所生成的序列数据。

503、基于多个序列数据通过词转化向量算法对模型进行训练, 获取更新的目标模型。

服务器基于多个序列数据通过词转化向量算法对模型进行训练, 获取更新的目标模型。

具体的, 通过词转化向量 word2vec 算法对多个序列数据进行分析, 以对模型进行训练, 15 更新模型。可通过输入层将多个序列数据中的词转换成向量的表示形式, 得到词向量, 隐层计算输入层输出的词向量的累加和, 得到一个总的向量, 总的向量与输出层的霍夫曼树的每一个非叶节点相连接, 根据语义关系和上下文关系对总的向量构建霍夫曼树, 通过能量函数对霍夫曼树进行概率计算。

可选的, 基于多个序列数据通过词转化向量算法对模型进行训练, 具体包括: 通过模 20 型根据多个序列数据创建霍夫曼树, 模型包括全连接层、多个连接层和归一化 Softmax 分类器; 通过全连接层将霍夫曼树中的各词进行分析, 获取多个预设长度的独热 one-hot 向量; 通过多个连接层对多个预设长度的 one-hot 向量进行梯度迭代处理, 获得列向量数据; 通过 Softmax 分类器对列向量数据进行处理, 获得词嵌入向量数据。

可选的, 获取词嵌入向量数据之后, 所述方法还包括: 通过 HashTable[MaxSize]函数 25 设置类型; 获取词嵌入向量数据的关键字以及关键字序列, 通过地址获取函数提取在散列表中查找、删除以及插入关键字的记录信息, 地址获取函数如下:

$H_i(\text{key}) = (\text{key} \% p + i) \% m, (1 \leq i \leq m-1)$, key 为关键字, m 为散列表长, i 为词嵌入向量数据的增量序列, p 为预设的素数; 根据类型和记录信息创建散列表, 初始化散列表, 将关键字序列插入散列表中, 并设置散列表的平均查找长度。通过将词嵌入向量数据中的关键 30 码值映射到散列表中的一个位置来访问记录, 以便于对历史案件的信息存储和提高对案件检索的效率。

可选的, 通过 Softmax 分类器对列向量数据进行处理之前, 方法还包括: 将列向量数据编码为有序的帧序列信息, 并对有序的帧序列信息设置定时器, 定时器包括定时时长大于有序的帧序列信息的发送往返时间的定时器; 设置自动重传请求 ARQ 协议与滑动窗口协 35 议, 以及设置连接 Softmax 分类器的连接层的发送滑动窗口, 并设置 Softmax 分类器的接收滑动窗口, 滑动窗口协议包括列向量数据的发送规则, 发送滑动窗口包括窗口大小为最大发送帧数的滑动窗口; 通过 ARQ 协议、滑动窗口协议、发送滑动窗口和接收滑动窗口将有序的帧序列信息传送至 Softmax 分类器, 以及通过 Softmax 分类器对连接层的连接层进行信息反馈。通过滑动窗口以对所述列向量数据的传输进行控制, 以免数据的堵塞和丢失, 40 同时也为了把历史列向量数据和当前列向量数据分离, 以使对当前列向量数据的处理速度不受因为历史列向量数据增加而减缓的影响, 从而提高案件检索系统对数据的处理效率。通过控制数据传送量, 以减压系统的缓存压力和提高传送效率。

504、获取目标文本信息, 并通过目标模型分析目标文本信息, 构建待检索知识图谱。

服务器获取目标文本信息, 并通过目标模型分析目标文本信息以构建待检索知识图谱。

具体的, 服务器对目标文本信息进行以清除文本信息中的噪声, 纠正不一致, 对目标 45 文本信息进行数据集成, 将目标文本信息有多个数据源合并成一个一致的数据存储, 对目

标文本信息进行聚集、删除冗余特征和聚类处理，对目标文本信息进行数据变换处理，以将目标文本信息压缩到较小的区间和机器能识别的表达式，以及对目标文本信息进行分词处理，对分词进行特征提取，获取特征信息，调用预置的卷积神经网络语言模型通过 NLP 算法将特征信息解释为知识图谱的结构化数据，结构化数据包括属性、属性对应的内容以及属性之间的关系，根据属性、属性对应的内容以及属性之间的关系创建待检索知识图谱。

505、在法律案件知识图谱中检索，获取与待检索知识图谱关联的案件信息和第一词嵌入向量数据，以及获取待检索知识图谱的第二词嵌入向量数据。

服务器在法律案件知识图谱中检索，获取与待检索知识图谱关联的案件信息和第一词嵌入向量数据，以及获取待检索知识图谱的第二词嵌入向量数据。

具体的，对待检索知识图谱进行特征提取，以及分析特征信息之间的联系，得到关联特征信息，关联特征信息包括案由信息和案件类型信息，根据关联特征信息遍历法律案件知识图谱，获取相同和/或相似的案件信息和案件信息对应的第一词嵌入向量数据，以及待检索知识图谱的第二词嵌入向量数据。以便于通过第一词嵌入向量数据和第二词嵌入向量数据计算待检索知识图谱的节点与法律案件知识图谱的节点之前的关系的相似度和内容的相似度。

506、根据第一词嵌入向量数据和第二词嵌入向量数据计算案件信息与目标文本信息之间的事实要素相似度、争议焦点相似度、证据相似度、争议焦点关联相似度、证据关联相似度和事实要素关联相似度。

服务器根据第一词嵌入向量数据和第二词嵌入向量数据计算案件信息与目标文本信息之间的事实要素相似度、争议焦点相似度、证据相似度、争议焦点关联相似度、证据关联相似度和事实要素关联相似度。

具体的，分别获取争论焦点、证据和事实要素的第一词嵌入向量数据，以及分别获取争论焦点、证据和事实要素的第二词嵌入向量数据，计算争论焦点的第一词嵌入向量数据和争论焦点的第二词嵌入向量数据的相似度得到争论焦点相似度，计算证据的第一词嵌入向量数据和证据的第二词嵌入向量数据的相似度得到证据相似度，计算事实要素的第一词嵌入向量数据和事实要素的第二词嵌入向量数据的相似度得到事实要素相似度。分别计算第一词嵌入向量数据与第二词嵌入向量数据的争议焦点关联相似度、证据关联相似度和事实要素关联相似度，分别得到案件信息与目标文本信息之间的争议焦点关联相似度、证据关联相似度和事实要素关联相似度。

507、根据第一预设权重比例计算事实要素相似度、争议焦点相似度和证据相似度的加权平均值，获得案件信息的第一相似度，以及根据第二预设权重比例计算争议焦点关联相似度、证据关联相似度和事实要素关联相似度的加权平均值，获得案件信息的第二相似度。

服务器根据第一预设权重比例计算事实要素相似度、争议焦点相似度和证据相似度的加权平均值，获得案件信息的第一相似度，以及根据第二预设权重比例计算争议焦点关联相似度、证据关联相似度和事实要素关联相似度的加权平均值，获得案件信息的第二相似度。

具体的，根据预设权重计算第一词嵌入向量数据和第二词嵌入向量数据的相似度加权平均值，得到第一相似度，其中，预设权重的设置方式为将争论焦点、证据和事实要素的权重依次递减。例如：争论焦点、证据和事实要素的权重分别为 40%、35%和 25%，在争论焦点方面上，计算得到的相似度为 0.5，在证据方面上，计算得到的相似度为 0.2，在事实要素方面上，计算得到的相似度为 0.3，则第一相似度=0.4*0.5+0.35*0.2+0.25*0.3=1.55。根据预设权重计算第一词嵌入向量数据和第二词嵌入向量数据的相似度加权平均值，得到第一相似度，其中，预设权重的设置方式为将争论焦点、证据和事实要素的权重依次递减。

508、识别并获取相似度大于第一预设阈值的第一相似度，以及识别并获取相似度大于第二预设阈值的第二相似度。

服务器识别并获取相似度大于第一预设阈值的第一相似度，以及识别并获取相似度大

于第二预设阈值的第二相似度。

具体的，判断第一相似度的相似度是否大于第一预设阈值，筛选出相似度大于第一预设阈值的第一相似度，判断第二相似度的相似度是否大于第二预设阈值，筛选出相似度大于第二预设阈值的第二相似度。通过对第一相似度和第二相似度筛选，以减少后续的计算操作，既能保证数据分析的质量和准确又能提高操作效率。

509、按照第三预设权重比例计算获取的第一相似度的加权平均值，获取目标第一相似度，以及按照第四预设权重比例计算获取的第二相似度的加权平均值，获取目标第二相似度。

服务器按照第三预设权重比例计算获取的第一相似度的加权平均值，获取目标第一相似度，以及按照第四预设权重比例计算获取的第二相似度的加权平均值，获取目标第二相似度。

具体的，在相似度大于第一预设阈值的第一相似度中按照第三预设权重比例计算加权平均值，得到一个总的相似度，即目标第一相似度。在相似度大于第二预设阈值的第二相似度中按照第四预设权重比例计算加权平均值，得到一个总的相似度，即目标第二相似度。第三预设权重可根据相似度大于第一预设阈值的第一相似度的相似度范围值进行设置权重，例如：相似度大于第一预设阈值的第一相似度为 0.2、0.35、0.15、0.55、0.5、0.27、0.87、0.9，则可设置第一相似度 0.15、0.2、0.27 的权重为 0.32，设置第一相似度 0.35、0.5、0.55 的权重为 0.3，设置第一相似度 0.87、0.9 的权重为 0.38，同理可得第四预设权重。上述举例内容仅作参考，其内容准确性和实际操作与否不作考虑。

510、将目标第一相似度和目标第二相似度相加，获取目标相似度。

服务器将目标第一相似度和目标第二相似度相加，获取目标相似度。通过将法律案件知识图谱中多个节点对应的相似度相加所得值作为法律案件知识图谱对应的法律案件与用户输入的待检索的案件的相似度，以提高检索的准确性。

511、按照目标相似度的值从高到低对案件信息进行排序，并输出经过排序的案件信息。

服务器按照目标相似度的值从高到低对案件信息进行排序，并输出经过排序的案件信息。

按照目标相似度的值从高到低对案件信息进行排序，并输出经过排序的案件信息，具体包括：将目标相似度按照相似度值从高到低对案件信息进行排序；接收用户输入的案件输出条件，根据案件输出条件确定目标案件信息，案件输出条件至少包括排序范围和相似度范围中的一个；对目标案件信息进行统计与分析，生成并输出可视化图表和文本信息。按照目标相似度的值从高至低对案件信息进行排序，得到排序的案件信息，对经过排序的案件进行解释说明，解释说明的内容包括案件信息与输入的文本信息的相似度、案件信息的判决意见和案件信息的证据提取方向等，并将案件信息和解释说明生成文本信息，以及对案件信息中各部分的相似度进行统计分析，生成可视化图表。当接收到用户输入的案件信息输出条件时，输出满足案件信息输出条件的案件信息的文本和可视化图表，案件输出条件包括但不限于案件信息的排序范围或案件的相似度范围。

上面对本申请实施例中基于知识图谱的案件检索方法进行了描述，下面对本申请实施例中基于知识图谱的案件检索装置进行描述，请参阅图 6，本申请实施例中基于知识图谱的案件检索装置的一个实施例包括：

第一构建单元 601，用于通过预置的模型分析文本信息构建法律案件知识图谱，并分析法律案件知识图谱构建节点集合数据，文本信息包括历史和/或实时的原告诉讼请求信息、被告的辩称信息、原告提供的证据信息、被告提供的证据信息和历史法律案件信息；

采样单元 602，用于将节点集合数据的各节点作为出发点，对节点集合数据进行随机游走采样，获得多个序列数据；

第一获取单元 603，用于基于多个序列数据通过词转化向量算法对模型进行训练，获取更新的目标模型；

第二构建单元 604, 用于获取目标文本信息, 并通过目标模型分析目标文本信息, 构建待检索知识图谱;

第二获取单元 605, 用于在法律案件知识图谱中检索, 获取与待检索知识图谱关联的案件信息和第一词嵌入向量数据, 以及获取待检索知识图谱的第二词嵌入向量数据;

- 5 计算单元 606, 用于根据第一词嵌入向量数据和第二词嵌入向量数据计算案件信息的第一相似度和第二相似度, 并将第一相似度和第二相似度相加, 获得目标相似度, 第一相似度指示案件信息在内容描述上的相似度, 第二相似度指示案件信息在逻辑关系上的相似度;

- 10 排序单元 607, 用于按照目标相似度的值从高到低对案件信息进行排序, 并输出经过排序的案件信息。

请参阅图 7, 本申请实施例中基于知识图谱的案件检索装置的另一个实施例包括:

第一构建单元 601, 用于通过预置的模型分析文本信息构建法律案件知识图谱, 并分析法律案件知识图谱构建节点集合数据, 文本信息包括历史和/或实时的原告告诉请信息、被告的辩称信息、原告提供的证据信息、被告提供的证据信息和历史法律案件信息;

- 15 采样单元 602, 用于将节点集合数据的各节点作为出发点, 对节点集合数据进行随机游走采样, 获得多个序列数据;

第一获取单元 603, 用于基于多个序列数据通过词转化向量算法对模型进行训练, 获取更新的目标模型;

- 20 第二构建单元 604, 用于获取目标文本信息, 并通过目标模型分析目标文本信息, 构建待检索知识图谱;

第二获取单元 605, 用于在法律案件知识图谱中检索, 获取与待检索知识图谱关联的案件信息和第一词嵌入向量数据, 以及获取待检索知识图谱的第二词嵌入向量数据;

- 25 第一计算单元 606, 用于根据第一词嵌入向量数据和第二词嵌入向量数据计算案件信息与目标文本信息之间的事实要素相似度、争议焦点相似度、证据相似度、争议焦点关联相似度、证据关联相似度和事实要素关联相似度;

第二计算单元 607, 用于根据第一预设权重比例计算事实要素相似度、争议焦点相似度和证据相似度的加权平均值, 获得案件信息的第一相似度, 以及根据第二预设权重比例计算争议焦点关联相似度、证据关联相似度和事实要素关联相似度的加权平均值, 获得案件信息的第二相似度;

- 30 第三获取单元 608, 用于识别并获取相似度大于第一预设阈值的第一相似度, 以及识别并获取相似度大于第二预设阈值的第二相似度;

第三计算单元 609, 用于按照第三预设权重比例计算获取的第一相似度的加权平均值, 获取目标第一相似度, 以及按照第四预设权重比例计算获取的第二相似度的加权平均值, 获取目标第二相似度;

- 35 第四计算单元 610, 用于将目标第一相似度和目标第二相似度相加, 获取目标相似度;

排序单元 611, 用于按照目标相似度的值从高到低对案件信息进行排序, 并输出经过排序的案件信息。

- 40 可选的, 第一构建单元 601 具体用于: 创建区块链结构网络, 获取区块链结构网络存储的文本信息, 区块链用于链接各个法律案件查询和/或管理平台的数据; 对文本信息进行词法分析、句法分析以及语义分析, 得到结构化数据, 并将结构化数据转换为标记语言数据; 对标记语言数据进行实体抽取以获取实体信息, 以及对标记语言数据进行关系抽取得到关系信息, 以及对标记语言数据进行属性抽取得到属性信息, 实体信息包括法律案件涉及的人名、组织/机构名、地理位置、事件/日期、字符值和金额值, 关系信息包括事实要素、人物、证据、争议焦点和法条中至少两者之间的关系; 对实体信息、关系信息和属性信息
- 45 信息进行信息融合处理, 以获取融合处理信息, 根据融合处理信息构建数据模型得到法律案件知识图谱, 并将法律案件知识图谱存储至区块链结构网络的关联数据库管理系统 MySQL

数据库。

可选的，采样单元 602 具体用于：将节点集合数据的各节点作为出发点，并获取出发点的下一节点的权重值；分析权重值以设置随机游走步数，以及分析出发点所在分支的节点数以设置游走阈值步数；当判断到随机游走步数小于游走阈值步数时，按照随机游走步数沿着连接出发点的下一节点的节点进行游走，并生成随机游走序列数据，获取多个序列数据；当判断到随机游走步数大于或等于游走阈值步数时，按照游走阈值步数沿着连接出发点的下一节点的节点进行游走，并生成随机游走序列数据，获取多个序列数据。

可选的，采样单元 602 还具体用于：以节点集合数据的每个节点作为出发点，获取除出发点之外的下游节点的权重值，根据权重值确定当前迭代次数；根据当前迭代次数，按照预设路径长度对节点集合数据进行随机游走；当检测到当前迭代次数小于预设迭代控制次数时，随机生成区间为 $[-1, 1]$ 之前的多维向量以及多维向量所在的多元函数；计算多元函数的函数值，若检测到函数值达到第一预设阈值，以及预设路径长度小于控制精度，则停止对节点集合数据随机游走，并获取在节点集合数据中进行随机游走时所生成的序列数据，控制精度用于停止对节点集合数据随机游走；若检测到函数值未达到第一预设阈值，以及预设路径长度不小于控制精度，则设置随机游走的行走步长为预设路径长度的一半，按照设置的随机游走的行走步长重新对节点集合数据随机游走，直至函数值达到第一预设阈值以及预设路径长度小于控制精度，并获取在节点集合数据中进行随机游走时所生成的序列数据。

可选的，第一获取单元 603 具体用于：通过模型根据多个序列数据创建霍夫曼树，模型包括全连接层、多个连接层和归一化 Softmax 分类器；通过全连接层将霍夫曼树中的各词进行分析，获取多个预设长度的独热 one-hot 向量；通过多个连接层对多个预设长度的 one-hot 向量进行梯度迭代处理，获得列向量数据；通过 Softmax 分类器对列向量数据进行处理，获得词嵌入向量数据。

可选的，第一获取单元具体 603 还具体用于：通过 HashTable[MaxSize] 函数设置类型；获取词嵌入向量数据的关键字以及关键字序列，通过地址获取函数提取在散列表中查找、删除以及插入关键字的记录信息，地址获取函数如下：

$H_i(\text{key}) = (\text{key} \% p + i) \% m, (1 \leq i \leq m-1)$ ，key 为关键字，m 为散列表长，i 为词嵌入向量数据的增量序列，p 为预设的素数；根据类型和记录信息创建散列表，初始化散列表，将关键字序列插入散列表中，并设置散列表的平均查找长度。

可选的，第一获取单元具体 603 还具体用于：将列向量数据编码为有序的帧序列信息，并对有序的帧序列信息设置定时器，定时器包括定时时长大于有序的帧序列信息的发送往返时间的定时器；设置自动重传请求 ARQ 协议与滑动窗口协议，以及设置连接 Softmax 分类器的连接层的发送滑动窗口，并设置 Softmax 分类器的接收滑动窗口，滑动窗口协议包括列向量数据的发送规则，发送滑动窗口包括窗口大小为最大发送帧数的滑动窗口；通过 ARQ 协议、滑动窗口协议、发送滑动窗口和接收滑动窗口将有序的帧序列信息传送到 Softmax 分类器，以及通过 Softmax 分类器对连接层的连接层进行信息反馈。

可选的，排序单元 611 具体用于：将目标相似度按照相似度值从高到低对案件信息进行排序；接收用户输入的案件输出条件，根据案件输出条件确定目标案件信息，案件输出条件至少包括排序范围和相似度范围中的一个；对目标案件信息进行统计与分析，生成并输出可视化图表和文本信息。

上面图 6 至图 7 从模块化功能实体的角度对本申请实施例中的基于知识图谱的案件检索装置进行详细描述，下面从硬件处理的角度对本申请实施例中基于知识图谱的案件检索设备进行详细描述。

图 8 是本申请实施例提供的一种基于知识图谱的案件检索设备的结构示意图，该基于知识图谱的案件检索设备 800 可因配置或性能不同而产生比较大的差异，可以包括一个或一个以上处理器（central processing units, CPU）801（例如，一个或一个以上处理器）

和存储器 809，一个或一个以上存储应用程序 807 或数据 806 的存储介质 808（例如一个或一个以上海量存储装置）。其中，存储器 809 和存储介质 808 可以是短暂存储或持久存储。存储在存储介质 808 的程序可以包括一个或一个以上模块（图示没标出），每个模块可以包括对签到管理设备中的一系列指令操作。更进一步地，处理器 801 可以设置为与存储介质 808 通信，在基于知识图谱的案件检索设备 800 上执行存储介质 808 中的一系列指令操作。

基于知识图谱的案件检索设备 800 还可以包括一个或一个以上电源 802，一个或一个以上有线或无线网络接口 803，一个或一个以上输入输出接口 804，和/或，一个或一个以上操作系统 805，例如 Windows Serve, Mac OS X, Unix, Linux, FreeBSD 等等。本领域技术人员可以理解，图 8 中示出的基于知识图谱的案件检索设备结构并不构成对基于知识图谱的案件检索设备的限定。处理器 801 可以执行上述实施例中第一构建单元 601、采样单元 602、第一获取单元 603、第二构建单元 604、第二获取单元 605、第一计算单元 606、第二计算单元 607、第三获取单元 608、第三计算单元 609、第四计算单元 610 和排序单元 611 的功能。

下面结合图 8 对基于知识图谱的案件检索设备的各个构成部件进行具体的介绍：

处理器 801 是基于知识图谱的案件检索设备的控制中心，可以按照基于知识图谱的案件检索方法进行处理。处理器 801 利用各种接口和线路连接整个基于知识图谱的案件检索设备的各个部分，通过运行或执行存储在存储器 809 内的软件程序和/或模块，以及调用存储在存储器 809 内的数据，执行基于知识图谱的案件检索设备的各种功能和处理数据，从而实现提高案件检索系统的可用性的功能。存储介质 808 和存储器 809 都是存储数据的载体，本申请实施例中，存储介质 808 可以是指储存容量较小，但速度快的内存储器，而存储器 809 可以是储存容量大，但储存速度慢的外存储器。

存储器 809 可用于存储软件程序以及模块，处理器 801 通过运行存储在存储器 809 的软件程序以及模块，从而执行基于知识图谱的案件检索设备 800 的各种功能应用以及数据处理。存储器 809 可主要包括存储程序区和存储数据区，其中，存储程序区可存储操作系统、至少一个功能所需的应用程序（通过预置的模型分析文本信息以构建法律案件知识图谱，并分析法律案件知识图谱以构建节点集合数据等）等；存储数据区可存储根据签到管理设备的使用所创建的数据（将节点集合数据的各节点作为出发点，对节点集合数据进行随机游走采样，获得多个序列数据等）等。此外，存储器 809 可以包括高速随机存取存储器，还可以包括非易失性存储器，例如至少一个磁盘存储器件、闪存器件、或其他非易失性固态存储器件。在本申请实施例中提供的基于知识图谱的案件检索方法程序和接收到的数据流存储在存储器中，当需要使用时，处理器 801 从存储器 809 中调用。

在计算机上加载和执行所述计算机程序指令时，全部或部分地产生按照本申请实施例所述的流程或功能。所述计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。所述计算机指令可以存储在计算机可读存储介质中，或者从一个计算机可读存储介质向另一计算机可读存储介质传输，其中所述计算机可读存储介质可以是非易失性，也可以是易失性。

权利要求书

1、一种基于知识图谱的案件检索方法，其中，包括：

通过预置的模型分析文本信息构建法律案件知识图谱，并分析所述法律案件知识图谱构建节点集合数据，所述文本信息包括历史和/或实时的原告诉请信息、被告的辩称信息、原告提供的证据信息、被告提供的证据信息和历史法律案件信息；

将所述节点集合数据的各节点作为出发点，对所述节点集合数据进行随机游走采样，获得多个序列数据；

基于多个所述序列数据通过词转化向量算法对所述模型进行训练，获取更新的目标模型；

获取目标文本信息，并通过所述目标模型分析所述目标文本信息，构建待检索知识图谱；

在所述法律案件知识图谱中检索，获取与所述待检索知识图谱关联的案件信息和第一词嵌入向量数据，以及获取所述待检索知识图谱的第二词嵌入向量数据；

根据所述第一词嵌入向量数据和所述第二词嵌入向量数据计算所述案件信息的第一相似度和第二相似度，并将所述第一相似度和所述第二相似度相加，获得目标相似度，所述第一相似度指示所述案件信息在内容描述上的相似度，所述第二相似度指示所述案件信息在逻辑关系上的相似度；

按照所述目标相似度的值从高到低对所述案件信息进行排序，并输出经过排序的案件信息。

2、根据权利要求1所述的基于知识图谱的案件检索方法，其中，所述将所述节点集合数据的各节点作为出发点，对所述节点集合数据进行随机游走采样，获得多个序列数据，包括：

将所述节点集合数据的各节点作为出发点，并获取所述出发点的下一节点的权重值；

分析所述权重值以设置随机游走步数，以及分析所述出发点所在分支的节点数以设置游走阈值步数；

当判断到所述随机游走步数小于所述游走阈值步数时，按照所述随机游走步数沿着连接所述出发点的下一节点的节点进行游走，并生成随机游走序列数据，获取多个序列数据；

当判断到所述随机游走步数大于或等于所述游走阈值步数时，按照所述游走阈值步数沿着连接所述出发点的下一节点的节点进行游走，并生成随机游走序列数据，获取多个序列数据。

3、根据权利要求2所述的基于知识图谱的案件检索方法，其中，所述基于多个所述序列数据通过词转化向量算法对所述模型进行训练，包括：

通过所述模型根据多个所述序列数据创建霍夫曼树，所述模型包括全连接层、多个连接层和归一化 Softmax 分类器；

通过全连接层将所述霍夫曼树中的各词进行分析，获取多个预设长度的独热 one-hot 向量；

通过多个所述连接层对多个所述预设长度的 one-hot 向量进行梯度迭代处理，获得列向量数据；

通过所述 Softmax 分类器对所述列向量数据进行处理，获得词嵌入向量数据。

4、根据权利要求3所述的基于知识图谱的案件检索方法，其中，所述通过多个所述连接层对多个所述预设长度的 one-hot 向量进行梯度迭代处理，获得列向量数据之后，所述通过所述 Softmax 分类器对所述列向量数据进行处理之前，所述方法还包括：

将所述列向量数据编码为有序的帧序列信息，并对所述有序的帧序列信息设置定时器，所述定时器包括定时时长大于所述有序的帧序列信息的发送往返时间的定时器；

设置自动重传请求 ARQ 协议与滑动窗口协议，以及设置连接所述 Softmax 分类器的连

接层的发送滑动窗口，并设置所述 Softmax 分类器的接收滑动窗口，所述滑动窗口协议包括所述列向量数据的发送规则，所述发送滑动窗口包括窗口大小为最大发送帧数的滑动窗口；

- 5 通过所述 ARQ 协议、所述滑动窗口协议、所述发送滑动窗口和所述接收滑动窗口将所述有序的帧序列信息传送至所述 Softmax 分类器，以及通过所述 Softmax 分类器对连接的连接层进行信息反馈。

5、根据权利要求 1 所述的基于知识图谱的案件检索方法，其中，所述根据所述第一词嵌入向量数据和所述第二词嵌入向量数据计算所述案件信息的第一相似度和第二相似度，包括：

- 10 根据所述第一词嵌入向量数据和所述第二词嵌入向量数据计算所述案件信息与所述目标文本信息之间的事实要素相似度、争议焦点相似度、证据相似度、争议焦点关联相似度、证据关联相似度和事实要素关联相似度；

- 根据第一预设权重比例计算所述事实要素相似度、所述争议焦点相似度和所述证据相似度的加权平均值，获得所述案件信息的第一相似度，以及根据第二预设权重比例计算所述争议焦点关联相似度、所述证据关联相似度和所述事实要素关联相似度的加权平均值，
15 获得所述案件信息的第二相似度。

6、根据权利要求 5 所述的基于知识图谱的案件检索方法，其中，所述将所述第一相似度和所述第二相似度相加，获得目标相似度，包括：

- 20 识别并获取相似度大于第一预设阈值的第一相似度，以及识别并获取相似度大于第二预设阈值的第二相似度；

按照第三预设权重比例计算获取的第一相似度的加权平均值，获取目标第一相似度，以及按照第四预设权重比例计算获取的第二相似度的加权平均值，获取目标第二相似度；

将所述目标第一相似度和所述目标第二相似度相加，获取目标相似度。

- 25 7、根据权利要求 1-6 中任意一项所述的基于知识图谱的案件检索方法，其中，所述按照所述目标相似度的值从高到低对所述案件信息进行排序，并输出经过排序的案件信息，包括：

将所述目标相似度按照相似度值从高到低对所述案件信息进行排序；

接收用户输入的案件输出条件，根据所述案件输出条件确定目标案件信息，所述案件输出条件至少包括排序范围和相似度范围中的一个；

- 30 对所述目标案件信息进行统计与分析，生成并输出可视化图表和文本信息。

8、一种基于知识图谱的案件检索装置，其中，所述装置包括：

第一构建单元，用于通过预置的模型分析文本信息构建法律案件知识图谱，并分析所述法律案件知识图谱构建节点集合数据，所述文本信息包括历史和/或实时的原告诉请信息、被告的辩称信息、原告提供的证据信息、被告提供的证据信息和历史法律案件信息；

- 35 采样单元，用于将所述节点集合数据的各节点作为出发点，对所述节点集合数据进行随机游走采样，获得多个序列数据；

第一获取单元，用于基于多个所述序列数据通过词转化向量算法对所述模型进行训练，获取更新的目标模型；

- 40 第二构建单元，用于获取目标文本信息，并通过所述目标模型分析所述目标文本信息，构建待检索知识图谱；

第二获取单元，用于在所述法律案件知识图谱中检索，获取与所述待检索知识图谱关联的案件信息和第一词嵌入向量数据，以及获取所述待检索知识图谱的第二词嵌入向量数据；

- 45 计算单元，用于根据所述第一词嵌入向量数据和所述第二词嵌入向量数据计算所述案件信息的第一相似度和第二相似度，并将所述第一相似度和所述第二相似度相加，获得目标相似度，所述第一相似度指示所述案件信息在内容描述上的相似度，所述第二相似度指

示所述案件信息在逻辑关系上的相似度；

排序单元，用于按照所述目标相似度的值从高到低对所述案件信息进行排序，并输出经过排序的案件信息。

- 5 9、一种基于知识图谱的案件检索设备，其中，包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现一种基于知识图谱的案件检索方法，所述方法包括：

通过预置的模型分析文本信息构建法律案件知识图谱，并分析所述法律案件知识图谱构建节点集合数据，所述文本信息包括历史和/或实时的原告诉请信息、被告的辩称信息、原告提供的证据信息、被告提供的证据信息和历史法律案件信息；

- 10 将所述节点集合数据的各节点作为出发点，对所述节点集合数据进行随机游走采样，获得多个序列数据；

基于多个所述序列数据通过词转化向量算法对所述模型进行训练，获取更新的目标模型；

- 15 获取目标文本信息，并通过所述目标模型分析所述目标文本信息，构建待检索知识图谱；

在所述法律案件知识图谱中检索，获取与所述待检索知识图谱关联的案件信息和第一词嵌入向量数据，以及获取所述待检索知识图谱的第二词嵌入向量数据；

- 20 根据所述第一词嵌入向量数据和所述第二词嵌入向量数据计算所述案件信息的第一相似度和第二相似度，并将所述第一相似度和所述第二相似度相加，获得目标相似度，所述目标相似度指示所述案件信息在内容描述上的相似度，所述第二相似度指示所述案件信息在逻辑关系上的相似度；

按照所述目标相似度的值从高到低对所述案件信息进行排序，并输出经过排序的案件信息。

- 25 10、根据权利要求 9 所述的基于知识图谱的案件检索设备，其中，所述将所述节点集合数据的各节点作为出发点，对所述节点集合数据进行随机游走采样，获得多个序列数据，包括：

将所述节点集合数据的各节点作为出发点，并获取所述出发点的下一节点的权重值；

分析所述权重值以设置随机游走步数，以及分析所述出发点所在分支的节点数以设置游走阈值步数；

- 30 当判断到所述随机游走步数小于所述游走阈值步数时，按照所述随机游走步数沿着连接所述出发点的下一节点的节点进行游走，并生成随机游走序列数据，获取多个序列数据；

当判断到所述随机游走步数大于或等于所述游走阈值步数时，按照所述游走阈值步数沿着连接所述出发点的下一节点的节点进行游走，并生成随机游走序列数据，获取多个序列数据。

- 35 11、根据权利要求 10 所述的基于知识图谱的案件检索设备，其中，所述基于多个所述序列数据通过词转化向量算法对所述模型进行训练，包括：

通过所述模型根据多个所述序列数据创建霍夫曼树，所述模型包括全连接层、多个连接层和归一化 Softmax 分类器；

- 40 通过全连接层将所述霍夫曼树中的各词进行分析，获取多个预设长度的独热 one-hot 向量；

通过多个所述连接层对多个所述预设长度的 one-hot 向量进行梯度迭代处理，获得列向量数据；

通过所述 Softmax 分类器对所述列向量数据进行处理，获得词嵌入向量数据。

- 45 12、根据权利要求 11 所述的基于知识图谱的案件检索设备，其中，所述通过多个所述连接层对多个所述预设长度的 one-hot 向量进行梯度迭代处理，获得列向量数据之后，所述通过所述 Softmax 分类器对所述列向量数据进行处理之前，所述方法还包括：

将所述列向量数据编码为有序的帧序列信息，并对所述有序的帧序列信息设置定时器，所述定时器包括定时时长大于所述有序的帧序列信息的发送往返时间的定时器；

5 设置自动重传请求 ARQ 协议与滑动窗口协议，以及设置连接所述 Softmax 分类器的连接层的发送滑动窗口，并设置所述 Softmax 分类器的接收滑动窗口，所述滑动窗口协议包括所述列向量数据的发送规则，所述发送滑动窗口包括窗口大小为最大发送帧数的滑动窗口；

通过所述 ARQ 协议、所述滑动窗口协议、所述发送滑动窗口和所述接收滑动窗口将所述有序的帧序列信息传送至所述 Softmax 分类器，以及通过所述 Softmax 分类器对连接

10 13、根据权利要求 9 所述的基于知识图谱的案件检索设备，其中，所述根据所述第一词嵌入向量数据和所述第二词嵌入向量数据计算所述案件信息的第一相似度和第二相似度，包括：

根据所述第一词嵌入向量数据和所述第二词嵌入向量数据计算所述案件信息与所述目标文本信息之间的事实要素相似度、争议焦点相似度、证据相似度、争议焦点关联相似度、
15 证据关联相似度和事实要素关联相似度；

根据第一预设权重比例计算所述事实要素相似度、所述争议焦点相似度和所述证据相似度的加权平均值，获得所述案件信息的第一相似度，以及根据第二预设权重比例计算所述争议焦点关联相似度、所述证据关联相似度和所述事实要素关联相似度的加权平均值，获得所述案件信息的第二相似度。

20 14、根据权利要求 13 所述的基于知识图谱的案件检索设备，其中，所述将所述第一相似度和所述第二相似度相加，获得目标相似度，包括：

识别并获取相似度大于第一预设阈值的第一相似度，以及识别并获取相似度大于第二预设阈值的第二相似度；

25 按照第三预设权重比例计算获取的第一相似度的加权平均值，获取目标第一相似度，以及按照第四预设权重比例计算获取的第二相似度的加权平均值，获取目标第二相似度；

将所述目标第一相似度和所述目标第二相似度相加，获取目标相似度。

15、根据权利要求 9-14 中任意一项所述的基于知识图谱的案件检索设备，其中，所述按照所述目标相似度的值从高到低对所述案件信息进行排序，并输出经过排序的案件信息，包括：

30 将所述目标相似度按照相似度值从高到低对所述案件信息进行排序；

接收用户输入的案件输出条件，根据所述案件输出条件确定目标案件信息，所述案件输出条件至少包括排序范围和相似度范围中的一个；

对所述目标案件信息进行统计与分析，生成并输出可视化图表和文本信息。

35 16、一种计算机可读存储介质，其中，包括指令，当所述指令在计算机上运行时，使得计算机执行

一种基于知识图谱的案件检索方法，所述方法包括：

通过预置的模型分析文本信息构建法律案件知识图谱，并分析所述法律案件知识图谱构建节点集合数据，所述文本信息包括历史和/或实时的原告诉请信息、被告的辩称信息、原告提供的证据信息、被告提供的证据信息和历史法律案件信息；

40 将所述节点集合数据的各节点作为出发点，对所述节点集合数据进行随机游走采样，获得多个序列数据；

基于多个所述序列数据通过词转化向量算法对所述模型进行训练，获取更新的目标模型；

45 获取目标文本信息，并通过所述目标模型分析所述目标文本信息，构建待检索知识图谱；

在所述法律案件知识图谱中检索，获取与所述待检索知识图谱关联的案件信息和第一

词嵌入向量数据, 以及获取所述待检索知识图谱的第二词嵌入向量数据;

根据所述第一词嵌入向量数据和所述第二词嵌入向量数据计算所述案件信息的第一相似度和第二相似度, 并将所述第一相似度和所述第二相似度相加, 获得目标相似度, 所述第一相似度指示所述案件信息在内容描述上的相似度, 所述第二相似度指示所述案件信息在逻辑关系上的相似度;

按照所述目标相似度的值从高到低对所述案件信息进行排序, 并输出经过排序的案件信息。

17、根据权利要求 16 所述的计算机可读存储介质, 其中, 所述将所述节点集合数据的各节点作为出发点, 对所述节点集合数据进行随机游走采样, 获得多个序列数据, 包括:

10 将所述节点集合数据的各节点作为出发点, 并获取所述出发点的下一节点的权重值;

分析所述权重值以设置随机游走步数, 以及分析所述出发点所在分支的节点数以设置游走阈值步数;

当判断到所述随机游走步数小于所述游走阈值步数时, 按照所述随机游走步数沿着连接所述出发点的下一节点的节点进行游走, 并生成随机游走序列数据, 获取多个序列数据;

15 当判断到所述随机游走步数大于或等于所述游走阈值步数时, 按照所述游走阈值步数沿着连接所述出发点的下一节点的节点进行游走, 并生成随机游走序列数据, 获取多个序列数据。

18、根据权利要求 17 所述的计算机可读存储介质, 其中, 所述基于多个所述序列数据通过词转化向量算法对所述模型进行训练, 包括:

20 通过所述模型根据多个所述序列数据创建霍夫曼树, 所述模型包括全连接层、多个连接层和归一化 Softmax 分类器;

通过全连接层将所述霍夫曼树中的各词进行分析, 获取多个预设长度的独热 one-hot 向量;

25 通过多个所述连接层对多个所述预设长度的 one-hot 向量进行梯度迭代处理, 获得列向量数据;

通过所述 Softmax 分类器对所述列向量数据进行处理, 获得词嵌入向量数据。

19、根据权利要求 18 所述的计算机可读存储介质, 其中, 所述通过多个所述连接层对多个所述预设长度的 one-hot 向量进行梯度迭代处理, 获得列向量数据之后, 所述通过所述 Softmax 分类器对所述列向量数据进行处理之前, 所述方法还包括:

30 将所述列向量数据编码为有序的帧序列信息, 并对所述有序的帧序列信息设置定时器, 所述定时器包括定时时长大于所述有序的帧序列信息的发送往返时间的定时器;

35 设置自动重传请求 ARQ 协议与滑动窗口协议, 以及设置连接所述 Softmax 分类器的连接层的发送滑动窗口, 并设置所述 Softmax 分类器的接收滑动窗口, 所述滑动窗口协议包括所述列向量数据的发送规则, 所述发送滑动窗口包括窗口大小为最大发送帧数的滑动窗口;

通过所述 ARQ 协议、所述滑动窗口协议、所述发送滑动窗口和所述接收滑动窗口将所述有序的帧序列信息传送至所述 Softmax 分类器, 以及通过所述 Softmax 分类器对连接层的连接层进行信息反馈。

40 20、根据权利要求 16 所述的计算机可读存储介质, 其中, 所述根据所述第一词嵌入向量数据和所述第二词嵌入向量数据计算所述案件信息的第一相似度和第二相似度, 包括:

根据所述第一词嵌入向量数据和所述第二词嵌入向量数据计算所述案件信息与所述目标文本信息之间的事实要素相似度、争议焦点相似度、证据相似度、争议焦点关联相似度、证据关联相似度和事实要素关联相似度;

45 根据第一预设权重比例计算所述事实要素相似度、所述争议焦点相似度和所述证据相似度的加权平均值, 获得所述案件信息的第一相似度, 以及根据第二预设权重比例计算所述争议焦点关联相似度、所述证据关联相似度和所述事实要素关联相似度的加权平均值,

获得所述案件信息的第二相似度。