



(12)发明专利申请

(10)申请公布号 CN 113326345 A

(43)申请公布日 2021.08.31

(21)申请号 202010130619.X

(22)申请日 2020.02.28

(71)申请人 拓尔思天行网安信息技术有限责任公司

地址 100084 北京市海淀区农大南路1号院2号楼2A201、202

(72)发明人 余江 王洪波 王亚强 张三海

(74)专利代理机构 北京志霖恒远知识产权代理事务所(普通合伙) 11435

代理人 郭栋梁

(51)Int.Cl.

G06F 16/33(2019.01)

G06F 16/36(2019.01)

G06F 40/295(2020.01)

G06F 40/30(2020.01)

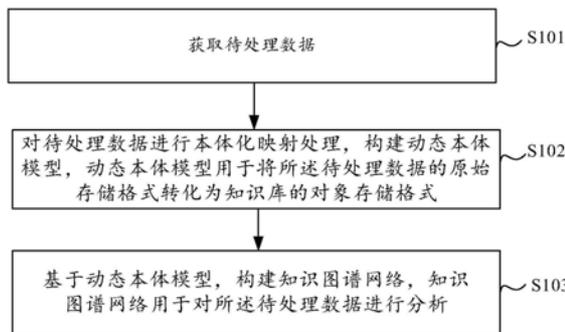
权利要求书2页 说明书11页 附图5页

(54)发明名称

基于动态本体的知识图谱分析、应用方法、平台及设备

(57)摘要

本申请公开了一种基于动态本体的知识图谱分析、应用方法、平台及设备,该方法包括:获取待处理数据;对待处理数据进行本体化映射处理,构建动态本体模型,所述动态本体模型用于将所述待处理数据的原始存储格式转化为知识库的对象存储格式;基于所述动态本体模型,构建关系图谱网络。该技术方案能够对海量的多源异构数据进行融合、关联、标注以及知识化处理,构建形成大规模领域关系的知识图谱网络,从而提高了数据深化应用能力,实现了数据的接入获取、存储、计算以及大规模的数据挖掘分析,进而能够实现从数据到知识到智能的升级转变,提高了对数据价值的挖掘效率。



1. 一种基于动态本体的知识图谱分析方法,其特征在于,包括:
获取待处理数据;
对所述待处理数据进行本体化映射处理,构建动态本体模型,所述动态本体模型用于将所述待处理数据的原始存储格式转化为知识库的对象存储格式;
基于所述动态本体模型,构建知识图谱网络,所述知识图谱网络用于对所述待处理数据进行分析。
2. 根据权利要求1所述的基于动态本体的知识图谱分析方法,其特征在于,对所述待处理数据进行映射处理,构建动态本体模型,包括:
采用实体对齐法对所述待处理数据进行文本语义分析,构建本体结构,所述本体结构包括对象、属性和关系;
对所述对象、所述属性和所述关系的描述进行分析处理,构建动态本体模型。
3. 根据权利要求1所述的基于动态本体的知识图谱分析方法,其特征在于,基于所述动态本体模型,构建知识图谱网络,包括:
根据所述动态本体模型,确定所述对象的属性联系、时空联系、语义联系和特征联系;
根据所述对象的属性联系、时空联系、语义联系和特征联系,构建知识图谱网络。
4. 根据权利要求1所述的基于动态本体的知识图谱分析方法,其特征在于,所述知识图谱网络包括实体与实体、实体与事件、实体与关系之间的联系。
5. 一种基于动态本体的知识图谱应用方法,其特征在于,所述方法包括:
获取知识图谱网络中的领域知识数据,所述知识图谱网络使用权利要求1-4任意一项所述的方法构建;
基于所述知识图谱网络的领域知识数据,采用可视化分析法进行知识数据应用、知识协同共享和多屏交互操作。
6. 根据权利要求5所述的基于动态本体的知识图谱应用方法,其特征在于,所述知识数据应用包括以下至少一项:知识检索、知识浏览、知识编辑、关系图分析、时空分析、对象统计分析和常用助手分析;所述知识协同共享包括以下至少一项:导出对象数据向其他用户进行共享、向其他用户分享关系图、获取其他用户分享的关系图、知识数据的订阅和发布。
7. 一种基于动态本体的知识图谱分析平台,其特征在于,所述平台包括:
获取模块,用于获取待处理数据;
模型构建模块,用于对所述待处理数据进行本体化映射处理,构建动态本体模型,所述动态本体模型用于将所述待处理数据的原始存储格式转化为知识库的对象存储格式;
网络构建模块,用于基于所述动态本体模型,构建知识图谱网络,所述知识图谱网络用于对所述待处理数据进行情报分析。
8. 一种基于动态本体的知识图谱应用平台,其特征在于,所述平台包括:
获取模块,用于获取知识图谱网络中的领域知识数据,所述知识图谱网络使用权利要求7所述的平台构建;
处理模块,用于基于所述知识图谱网络的领域知识数据,采用可视化分析法进行知识数据应用、知识协同共享和多屏交互操作。
9. 一种计算机设备,其特征在于,所述终端设备包括存储器、处理器以及存储在存储器中并可在处理器上运行的计算机程序,所述处理器用于执行所述程序时实现如权利要求1-

4或5-6任一项所述的方法。

10. 一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序用于实现如权利要求1-4或5-6任一项所述的方法。

基于动态本体的知识图谱分析、应用方法、平台及设备

技术领域

[0001] 本发明一般涉及数据处理技术领域,具体涉及一种基于动态本体的知识图谱分析、应用方法、平台及设备。

背景技术

[0002] 随着互联网技术的快速发展,产生了海量的多源异构数据,在各类智能化应用需求不断涌现的过程中,尤其是当面对社会公共突发事件时,需要通过大数据分析以跟踪目标人物的活动轨迹和密切接触人员行踪等相关数据信息。因此,为了满足人工智能对海量多源异构数据及数据关联挖掘的需要,对海量异构数据进行综合处理分析显得十分重要。

[0003] 目前,对海量异构数据的分析工具比较简单,无法直观可视化地得到数据的分析结果,且对专业应用数据分析不够深入,导致对数据价值的挖掘效率低下,难以支撑人工智能对海量多源异构数据及数据关联挖掘的需要。

发明内容

[0004] 鉴于现有技术中的上述缺陷或不足,期望提供一种基于动态本体的知识图谱分析、应用方法、平台及设备,能够通过构建知识图谱网络,快速实现从数据到知识的转化,很大程度上提高了对数据价值的挖掘效率。

[0005] 第一方面,本申请实施例提供了一种基于动态本体的知识图谱分析方法,该方法包括:

[0006] 获取待处理数据;

[0007] 对所述待处理数据进行本体化映射处理,构建动态本体模型,所述动态本体模型用于将所述待处理数据的原始存储格式转化为知识库的对象存储格式;

[0008] 基于所述动态本体模型,构建知识图谱网络,所述知识图谱网络用于对所述待处理数据进行处理分析。

[0009] 在其中一个实施例中,对所述待处理数据进行映射处理,构建动态本体模型,包括:

[0010] 采用实体对齐法对所述待处理数据进行文本语义分析,构建本体结构,所述本体结构包括对象、属性和关系;

[0011] 对所述对象、所述属性和所述关系的描述进行分析处理,构建动态本体模型。

[0012] 在其中一个实施例中,基于所述动态本体模型,构建知识图谱网络,包括:

[0013] 根据所述动态本体模型,确定所述对象的属性联系、时空联系、语义联系和特征联系;

[0014] 根据所述对象的属性联系、时空联系、语义联系和特征联系,构建关系图谱网络。

[0015] 在其中一个实施例中,所述知识图谱网络包括实体与实体、实体与事件、实体与关系之间的联系。

[0016] 第二方面,本申请实施例提供了一种基于动态本体的知识图谱应用方法,该方法

包括：

[0017] 获取知识图谱网络中的领域知识数据；

[0018] 基于所述知识图谱网络的领域知识数据，采用可视化分析法进行知识数据应用、知识协同共享和多屏交互操作。

[0019] 在其中一个实施例中，所述知识数据应用包括以下至少一项：知识检索、知识浏览、知识编辑、关系图分析、时空分析、对象统计分析和常用助手分析；所述知识协同共享包括以下至少一项：导出对象数据向其他用户进行共享、向其他用户分享关系图、获取其他用户分享的关系图、知识数据的订阅和发布。

[0020] 第三方面，本申请提供了一种基于动态本体的知识图谱分析平台，该平台包括：

[0021] 获取模块，用于获取待处理数据；

[0022] 模型构建模块，用于对所述待处理数据进行本体化映射处理，构建动态本体模型，所述动态本体模型用于将所述待处理数据的原始存储格式转化为知识库的对象存储格式；

[0023] 网络构建模块，用于基于所述动态本体模型，构建知识图谱网络，所述知识图谱网络用于对所述待处理数据进行处理分析。

[0024] 第四方面，本申请提供了一种基于动态本体的知识图谱应用平台，该平台包括：

[0025] 获取模块，用于获取知识图谱网络中的领域知识数据；

[0026] 处理模块，用于基于所述知识图谱网络的领域知识数据，采用可视化分析法进行知识数据应用、知识协同共享和多屏交互操作。

[0027] 第五方面，本申请实施例提供一种设备，包括存储器、处理器以及存储在存储器上并可在处理器上运行的计算机程序，该处理器执行该程序时实现如上述第一方面或第二方面的方法。

[0028] 第六方面，本申请实施例提供一种计算机可读存储介质，其上存储有计算机程序，该计算机程序用于实现如上第一方面或第二方面的方法。

[0029] 本申请实施例中提供的基于动态本体的知识图谱分析、应用方法、平台及设备，通过获取待处理数据，并对待处理数据进行本体化映射处理，构建动态本体模型，该动态本体模型用于将待处理数据的原始存储方式转化为知识库的对象存储方式，并基于动态本体模型，构建知识图谱网络，以使得通过知识图谱网络对待处理数据进行情报分析。该技术方案能够对海量的多源异构数据进行融合、关联、标注以及知识化处理，构建形成大规模领域关系图谱网络，从而提高了数据深化应用能力，实现了数据的接入获取、存储、计算以及大规模的数据挖掘分析，进而能够实现从数据到知识到智能的升级转变，提高了对数据价值的挖掘效率，能够有效支撑各领域信息化向智能化应用的发展。

附图说明

[0030] 通过阅读参照以下附图所作的对非限制性实施例所作的详细描述，本申请的其它特征、目的和优点将会变得更明显：

[0031] 图1为本申请实施例提供的基于动态本体的知识图谱分析方法的实施环境架构图；

[0032] 图2为本申请实施例提供的构建动态本体模型方法的流程示意图；

[0033] 图3为本申请实施例提供的基于动态本体的知识图谱分析方法的流程示意图；

- [0034] 图4为本申请实施例提供的基于动态本体的知识图谱应用方法的流程示意图；
- [0035] 图5为本申请实施例提供的基于动态本体的知识图谱分析平台的流程示意图；
- [0036] 图6为本申请另一实施例提供的基于动态本体的知识图谱分析平台的结构示意图；
- [0037] 图7为本申请实施例提供的基于动态本体的知识图谱应用平台的结构示意图；
- [0038] 图8为本申请实施例提供的基于动态本体的知识图谱分析平台的架构示意图；
- [0039] 图9为本申请实施例示出的一种计算机设备的结构示意图。

具体实施方式

[0040] 下面结合附图和实施例对本申请作进一步的详细说明。可以理解的是，此处所描述的具体实施例仅仅用于解释相关发明，而非对该发明的限定。另外还需要说明的是，为了便于描述，附图中仅示出了与发明相关的部分。

[0041] 需要说明的是，在不冲突的情况下，本申请中的实施例及实施例中的特征可以相互组合。下面将参考附图并结合实施例来详细说明本申请。

[0042] 如背景技术中提到的，人工智能已逐渐成为未来各领域的竞争焦点，各行各业纷纷踏上智能化升级与转型的道路，各类智能化应用需求不断涌现，为了满足人工智能对海量多源异构数据及数据关联挖掘的需要，对大数据进行综合分析至关重要。现有的相关技术中对海量异构数据的分析工具比较简单，且对数据管理模式比较单一，无法直观可视化地得到数据的分析结果，且对专业应用数据分析不够深入，导致对数据价值挖掘不够，难以支撑人工智能对海量多源异构数据及数据关联挖掘的需要。

[0043] 基于以上缺陷，本申请提供了一种基于动态本体的知识图谱分析及应用方法，能够通过对海量的多源异构数据进行融合、关联、标注以及知识化处理，构建形成大规模领域关系图谱网络，实现了数据的接入获取、存储、计算以及大规模数据挖掘分析，快速实现从数据到知识的转化，很大程度上提高了对数据价值的挖掘效率，能够有效支撑各领域信息化向智能化应用的发展。

[0044] 本申请提供的基于动态本体的知识图谱分析方法，能够支撑多源异构数据融合处理、语义检索、情报分析、执法调查、辅助决策研究和开源情报分析等，可以应用于情报机构、调查机构、政府、媒体等不同领域机构，为国家安全、公共安全、政府决策和企业运营提供从多源异构数据融合、存储、知识化处理、图谱关联挖掘分析、地理空间分析、时序分析和可视化展现等分析场景的全面支持。

[0045] 图1是本申请实施例提供的一种基于动态本体的知识图谱分析方法的实施环境架构图。如图1所示，该实施环境架构包括：终端100和服务器200。

[0046] 终端100可以是各类AI应用场景中的终端设备。例如，终端100可以是智能电视、智能电视机顶盒等智能家居设备，或者终端100可以是智能手机、平板电脑以及电子书阅读器等移动式便携终端，或者，该终端100可以是智能眼镜、智能手表等智能可穿戴设备，本实施例对此不进行具体限定。

[0047] 其中，终端100中可安装有基于自然语言处理的AI应用。比如，该AI应用可以是智能搜索、智能问答等应用。

[0048] 服务器200可以是一台服务器，也可以是由若干台服务器构成的服务器集群，或者

服务器200可以包含一个或多个虚拟化平台,或者服务器200可以是一个云计算服务中心。

[0049] 其中,服务器200可以是为上述终端100中安装的AI应用提供后台服务的服务器设备。

[0050] 终端100与服务器200之间通过有线或无线网络建立通信连接。可选的,上述的无线网络或有线网络使用标准通信技术和/或协议。网络通常为因特网、但也可以是任何网络,包括但不限于局域网(Local Area Network,LAN)、城域网(Metropolitan Area Network,MAN)、广域网(Wide Area Network,WAN)、移动、有线或者无线网络、专用网络或者虚拟专用网络的任何组合。

[0051] 为了便于理解和说明,下面通过图2至图9详细阐述本申请实施例提供的基于动态本体的知识图谱分析、应用方法、平台及设备。

[0052] 图2所示为本申请实施例的基于动态本体的知识图谱分析方法的流程示意图,该方法可以由计算机设备执行,该计算机设备可以是上述图1所示系统中的服务器200或者终端100,或者,该计算机设备也可以是终端100和服务器200的结合。如图2所示,该方法包括:

[0053] S101、获取待处理数据。

[0054] 具体的,计算机设备上可以运行有知识图谱分析系统,该分析系统支持对接各类数据库数据资源,可一键导入不同类型的数据库,如Oracle、MPP、HBASE等,同时提供数据批量导入功能,可以将各种渠道中收集到的数据方便地导入到系统中。上述待处理数据可以通过客户端导入的,也可以通过网页导入的,其中,通过客户端导入待处理数据的形式可以从关系数据库中导入数据,该关系数据库可以是TRS水晶分布式数据库,也可以是从excel格式的表格中导入,还可以是对特定格式或特定数据进行导入。通过网页导入待处理数据指的是通过网页的方式将待处理数据导入至平台中,例如可以通过终端设备内部或者外部搜索引擎,该搜索引擎例如可以是TRS海贝分布式搜索引擎,将搜索引擎返回的网页内容导入至终端设备的系统平台中,还可以是通过接口的方式将浏览器内展示的内容导入至终端设备的系统平台中。

[0055] 可选的,该待处理数据可以是结构化数据,也可以是非结构化数据,还可以是半结构化数据。其中,结构化数据是指由二维表结构来逻辑表达和实现的数据,通过关系数据库进行存储和管理,如关系数据库里的表;非结构化数据是数据结构不规则或不完整,没有预定义的数据模型,包括所有格式的办公文档、文本、图片、HTML、各类报表、图像和音频/视频信息等。

[0056] S102、对待处理数据进行本体化映射处理,构建动态本体模型,该动态本体模型用于将待处理数据的原始存储方式转化为知识库的对象存储方式。

[0057] 具体的,待处理数据的原始存储结构可以是表结构、数据库等存储方式;知识库的对象存储方式是一种将数据作为对象进行管理的计算机数据存储体系结构,包括对象、属性和关系,用于描述现实世界中的各种事物。其中,对象可以包括实体、事件、文档和多媒体等,实体可以是人、位置等,属性可以是对象的性质,如人的身份证、身高、体重等信息,关系可以是对象的响应,即该实体与其他实体之间的关联事件。动态本体模型是通过动态本体技术构建出的模型,可以将待处理数据的表结构、数据库等存储方式进行本体化映射处理后转化为知识库的对象存储方式。

[0058] 在通过各类数据库资源或其他渠道获取到待处理数据后,可以先获取待处理数据

的类别信息,该类别信息可以是结构化数据、非结构化数据、半结构化数据,基于不同数据的类别信息,构建动态本体模型。

[0059] 示例性地,例如获取到的导入数据为通联数据、交易数据、人的属性信息等数据,经过对该数据进行映射处理后,可以是以人为对象进行存储,分别将人的属性、人的关系进行集中存储,使得在发现实体与实体之间关系的时候,可以很快的搜索或关联到该实体与其他实体之间的关联关系,如人与其他人之间的关系,人与某个地点之间的关系等。

[0060] 可选的,图2为本申请实施例提供的动态本体模型构建方法的流程示意图,可以参见图2所示,步骤102可以包括如下步骤:

[0061] S1021、采用实体对齐法对待处理数据进行文本语义分析,构建本体结构,该数据结构包括对象、属性和关系。

[0062] S1022、将对象、属性和关系的描述进行分析处理,构建动态本体模型。

[0063] 需要说明的是,在获取到接入的待处理数据后,该待处理数据包括结构化数据、非结构化数据和半结构化数据,通过对结构化数据、半结构化数据和非结构化数据关系挖掘,数据分析、文本语义分析等,抽取出对象、属性、关系等本体结构,可以使用多源数据的实体对齐技术对待处理数据进行处理,其中实体对齐也称为实体匹配或实体解析或者实体链接。并通过对象的描述来确定对象具有的属性和关系来概念化本体,将对象和属性进行建模,还可以定义和映射两个对象之间的连接是关系还是事件,将关系和事件进行建模。

[0064] 其中,结构化数据导入的时候可以映射对象和表属性之间的关系,建立动态本体模型。

[0065] 可选的,可以使用基于逻辑的知识推理方法进行本体构建,该基于逻辑的推理方式主要包括一阶谓词逻辑、描述逻辑以及规则等。其中,一阶位于逻辑推理是以命题为基本进行推理,而命题又包括个体和谓词,逻辑中的个体对应关系图谱网络中的实体对象,具有客观独立性,可以是具体一个或泛指一类;谓词则描述了个体的性质和个体间的关系,从而实现基于谓词、描述逻辑和规则的知识推理。

[0066] 本实施例中基于动态本体的知识对象建模,融合动态本体理论技术,能够对多源异构的待处理数据进行本体化标记处理,进而重新组织构建数据结构,使得数据能够从多源的数据格式,被转换映射为统一的数据对象,实现了数据的整合处理,并采用多源数据的实体对齐技术消除了异构数据中实体冲突、指向不明等不一致问题,基于动态本体技术在知识图谱顶层采用概率模型进行全局集体实体对齐,从而帮助机器理解多源异构的数据,进而形成高质量的知识。

[0067] S103、基于动态本体模型,构建知识图谱网络,该知识图谱网络用于对待处理数据进行处理分析。

[0068] 具体的,上述知识图谱网络是一种基于图的数据结构,由节点和边组成,每个节点表示现实世界中存在的“实体”,每条边为实体与实体之间的“关系”。通俗来讲,关系图谱网络就是基于知识图谱技术构建的把所有不同种类的信息连接在一起而得到的一个关系网络,提供了从“关系”的角度去分析问题的能力。其中,知识图谱称为知识领域可视化或知识领域映射地图,是展示知识发展进程与结构关系的一系列各种不同的图形,用可视化技术描述知识资源及其载体,挖掘、分析、构建、绘制和显示知识与它们之间的联系。具体来说,知识图谱是通过将应用数学、图形学、信息可视化技术、信息科学等学科的理论与方法与计

量学引文分析、共现分析等方法的结合,并利用可视化的图谱形象地展示学科的核心结构、发展历史、前沿领域以及整体知识架构达到多学科融合目的现代理论。它能够把复杂的知识领域通过数据挖掘、信息处理、知识计量和图形绘制而显示出来。

[0069] 在构建出动态本体模型之后,可以确定对象的属性联系、时空联系、语义联系和特征联系,并根据对象的属性联系、时空联系、语义联系和特征联系,构建知识图谱网络。其中,知识图谱网络包括实体与实体、实体与事件、实体与关系之间的联系。

[0070] 构建出的知识图谱网络中的数据可以存储在领域知识资源库中,可以进行本体管理和知识版本管理,本体管理可以是对知识资源库中的本体对象、属性的描述、关系等进行构建及管理,支持对象类型、属性类型、关系类型的用户自定义,即用户能够对本体类型、属性类型、关系类型对应的值进行修改。知识版本管理可以是版本控制能让用户知道数据或本体的变化带来的影响,同时知识可通过人工编辑、自动回标等方法进行内容更新,并具有版本的比对、回退、编辑等功能。

[0071] 需要说明的是,由于不同的组织机构眼中的世界是不同的,所以构建的动态本体模型也会随着事件的变化而变化,随着数据源的添加、删除和重新构造而更新。

[0072] 本申请实施例中提供的基于动态本体的知识图谱分析方法,通过获取待处理数据,并对待处理数据进行本体化映射处理,构建动态本体模型,该动态本体模型用于将待处理数据的原始存储方式转化为知识库的对象存储方式,并基于动态本体模型,构建关系图谱网络,以使得通过关系图谱网络对待处理数据进行处理分析。该技术方案能够对海量的多源异构数据进行融合、关联、标注以及知识化处理,构建形成大规模领域关系图谱网络,从而提高了数据深化应用能力,实现了数据的接入获取、存储、计算以及大规模的数据挖掘分析,进而能够实现从数据到知识到智能的升级转变,提高了对数据价值的挖掘效率,能够有效支撑各领域信息化向智能化应用的发展。

[0073] 进一步地,在上述实施例的基础上,如图3所示,该方法包括以下步骤:

[0074] S201、获取待处理数据。

[0075] S202、对待处理数据进行本体化映射处理,构建动态本体模型,动态本体模型用于将待处理数据的原始存储格式转化为知识库的对象存储格式。

[0076] S203、基于动态本体模型,构建知识图谱网络,知识图谱网络用于对待处理数据进行处理分析。

[0077] S204、基于知识图谱网络,采用可视化分析法进行知识数据应用、知识协同共享和多屏交互操作。

[0078] 具体的,计算机设备在获取到接入的待处理数据后,可以对待处理数据进行文本语义分析,抽取出对象、属性、关系等数据结构,通过对象的描述来确定对象具有的属性和关系来概念化本体,将对象和属性进行建模,还可以定义和映射两个对象之间的连接是关系还是事件,将关系和事件进行建模。

[0079] 在构建出动态本体模型之后,可以确定对象的属性联系、时空联系、语义联系和特征联系,构建知识图谱网络,并基于知识图谱网络的领域知识数据,采用可视化分析方法进行知识数据应用、知识协同共享和多屏交互操作等可视化操作处理。其中,可视化分析技术是通过图像的方式对信息元之间的关系进行描述和展现,并运用关联分析、网络分析、路径分析、时序分析、空间分析等多种图形分析方式来发现和揭示数据中隐含的信息和关联。

[0080] 可选的,基于知识图谱网络进行知识数据应用包括进行知识的关联检索、知识浏览、知识编辑、关系图分析、时空分析、对象统计分析和常用助手分析等。

[0081] 其中,知识检索的搜索功能类似百度和Google的检索,能够提供简单检索、高级检索、快速检索和过滤器高级检索功能,主要是针对内部知识库资源的检索。知识浏览指的是系统提供统一的知识浏览样式,支持对系统知识中的实体、事件、文档等对象的浏览。知识编辑指的是系统提供统一的知识编辑功能,能够支持对系统知识中的实体、事件、文档等对象的编辑。

[0082] 需要说明的是,上述关联图分析中的关联图是将本体结构中的实体、时间、文档等作为节点和关系作为边进行建模得到的。关联图应用通常是分析人员在分析过程中使用最多的应用,关联图上能够根据所选对象进行上下文搜索,调整布局样式以及结合各种助手完成复杂操作。关联图分析能够支持展示多种布局方式,如自由布局、网格布局、圆环布局、线性布局、分层布局,且支持关联图与地图互联操作,支持互动操作和快照功能等,快照支持导出PPT,并能够回退到任意历史快照。关联图每次支持一万个对象上图分析,最大支持十万个对象上图分析。关联图支持多屏数据联动功能,支持多个屏幕运行实体浏览、关联图、地图互动显示等,并能够实现数据显示互动。

[0083] 时空分析指的是通过时间轴、地图图层进行地图标注、地址对象解析、地图热力图分析等地图分析工作,支持通过时间轴的方式组织数据集,并进行联动分析展现。时间轴支持设置事件粒度、事件区间框选、事件窗口拖拽等方式切换选中数据集,以支撑时间维度的分析作业。

[0084] 对象统计分析支持将当前调查的所有对象、属性进行统计分析,支持图表、属性值直方图、分组统计、链接类型直方图、属性统计图和时间图等,且支持对任意数据和属性进行下钻,其中下钻指的是从当前数据往下展开至下一层数据。

[0085] 同样,系统还提供多种常用的分析助手,分析助手可与关系图、地图进行联动分析。

[0086] 进一步的,在构建出知识图谱网络之后,可以基于知识图谱网络进行知识协同共享,例如导出对象数据向其他用户进行共享,对象数据导出支持在关联图上选中数据,快速导出excel或word,还可以导出数据的属性等;还支持将系统中分析的关系图分享给其他用户,关系图分享支持根据权限进行用户分享,用户可以在操作界面上选择使用模板进行分享或分享给特定用户或用户组;还可以对分享的关系图进行命名和添加描述信息。

[0087] 另外,系统还支持获取其他用户分享的关系图,获取时用户可以通过在操作界面上快速找到要获取的关系图,支持设置关系图创建者、开始时间、结束时间等条件进行检索,找到分享的关系图后可进行预览,并添加至用户本体的关系图中。系统还可以根据用户操作需要,能够保存关联图的每一次操作、布局的快照,并能够导出图片、PPT等,并且系统保存的每一张快照都可以快速浏览和回退。系统还可以提供知识订阅和发布功能,支持用户订阅相关的知识数据、发布自有的知识等。

[0088] 进一步地,系统还可以支持用户的多屏联动分析,用户分析过程可将不同的分析内容、模块分散到多个屏幕进行独立分析,同时支持对标注、编辑、发布等用户操作行为的协同处理。

[0089] 本实施例提供中基于知识图谱网络的可视化分析方法能够帮助用户将大量的、未

知的、低关联性的、低价值的信息转化为少量的、易于理解的、高关联性的、高价值的可操作知识,为数据分析、数据挖掘工作提供很大的帮助,提高了数据处理的效率。

[0090] 本申请实施例中数据从多源的数据格式、被转换映射为统一的数据对象,从而关联现实世界中的人、地点、事物、事件、以及之间的关系,极大的简化了数据集成的流程,缩短了数据整合导入和转换的时间,并且能够通过分布式计算和存储架构,处理PB级数据,包括业务数据、海量日志、网络流浪、交易数据等结构化数据和文档、邮件、网页等非结构化数据。能够通过基于关系图谱网络支持海量外部非结构化数据的检索以及对接查询外部的结构化数据,从而在多源、异构、多模态条件下实现全文索引和搜索,实现对现实对象和实体的转换、映射和关联等操作,进而使得不同用户之间实现数据共享,极大的挖掘到数据价值,提高了数据利用率。

[0091] 另一方面,图5为本申请实施例提供的一种基于动态本体的知识图谱分析平台的结构示意图。该系统可以为终端或服务器内的装置,如图5所示,该平台500包括:

[0092] 获取模块510,用于获取待处理数据;

[0093] 模型构建模块520,用于对待处理数据进行本体化映射处理,构建动态本体模型,动态本体模型用于将待处理数据的原始存储格式转化为知识库的对象存储格式;

[0094] 网络构建模块530,用于基于动态本体模型,构建知识图谱网络,知识图谱网络用于对待处理数据进行分析。

[0095] 可选的,参见图6,模型构建模块520,包括:

[0096] 第一构建单元521,用于采用实体对齐法对待处理数据进行文本语义分析,构建本体结构,本体结构包括对象、属性和关系;

[0097] 第二构建单元522,用于对对象、属性和关系的描述进行分析处理,构建动态本体模型。

[0098] 可选的,网络构建模块530,包括:

[0099] 确定单元531,用于根据动态本体模型,确定所述对象的属性联系、时空联系、语义联系和特征联系;

[0100] 第三构建单元532,用于根据对象的属性联系、时空联系、语义联系和特征联系,构建知识图谱网络。

[0101] 可选的,知识图谱网络包括实体与实体、实体与事件、实体与关系之间的联系。

[0102] 另一方面,本申请提供了一种基于动态本体的知识图谱应用平台的结构示意图,参见图7,该平台包括:

[0103] 获取模块540,用于获取知识图谱网络中的领域知识数据;

[0104] 处理模块550,用于基于知识图谱网络的领域知识数据,采用可视化分析法进行知识数据应用、知识协同共享和多屏交互操作。

[0105] 可选的,知识应用包括以下至少一项:知识检索、知识浏览、知识编辑、关系图分析、时空分析、对象统计分析和常用助手分析;知识协同共享包括以下至少一项:导出对象数据向其他用户进行共享、向其他用户分享关系图、获取其他用户分享的关系图、知识数据的订阅和发布。

[0106] 请参见图8所示,该平台可以是通过从大数据资源库导入数据,例如通过分布式数据库、Hbase开源数据库,HDFS文件系统,分布式检索引擎、通过互联网采集的方式、或者通

过文本挖掘的方式导入待处理数据,可以使用开源ETL工具对待处理数据进行数据清洗、转换,并加载至目标端;该系统采用了大数据检索技术、动态本体技术、关系图、地图分析、分布事务等应用技术,从而能够基于关系图谱网络实现多种应用功能,例如数据整合、知识检索、知识浏览、知识编辑、知识管理、关系图分析、时空分析、对象统计分析、常用助手分析等。

[0107] 可以理解的是,本实施例的基于动态本体的知识图谱分析平台的各功能模块的功能可根据上述方法实施例中的方法具体实现,其具体实现过程可以参照上述方法实施例的相关描述,在此不再赘述。

[0108] 本实施例提供的基于动态本体的知识图谱分析平台建立在信息化统一管理平台、统一安全体系的框架下,基于云计算大数据架构,开展水晶球分析建设。引入多种应用技术,建设基于本体论的数据建模功能,包括领域知识库建设、专题分析库建设、知识检索和应用分析展现功能等。能够实现数据的接入、存储、计算以及大规模的数据挖掘分析,为系统分析服务提供支撑,应该功能主要包括本体管理、知识抽取、知识管理、知识检索、应用分析展现等,在应用功能的基础上,支持用户进行综合分析、专案分析、辅助决策、沙盘推演等各类专项分析或研判工作。

[0109] 综上所述,本申请实施例中提供的基于动态本体的知识图谱分析平台,通过获取模块获取待处理数据,并通过模型构建模块对待处理数据进行本体化映射处理,构建动态本体模型,该动态本体模型用于将待处理数据的原始存储方式转化为知识库的对象存储方式,并使用网络构建模块基于动态本体模型,构建知识图谱网络,以使得通过知识图谱网络对待处理数据进行处理分析。该技术方案能够对海量的多源异构数据进行融合、关联、标注以及知识化处理,构建形成大规模领域关系图谱网络,从而提高了数据深化应用能力,实现了数据的接入获取、存储、计算以及大规模的数据挖掘分析,进而能够实现从数据到知识到智能的升级转变,提高了对数据价值的挖掘效率,能够有效支撑各领域信息化向智能化应用的发展。

[0110] 另一方面,本申请实施例提供的设备,包括存储器、处理器以及存储在存储器上并可在处理器上运行的计算机程序,该处理器执行该程序时实现如上述的基于动态本体的知识图谱分析方法及应用方法。

[0111] 下面参考图9,图9为本申请实施例的终端设备的计算机系统的结构示意图。

[0112] 如图9所示,计算机系统300包括中央处理单元(CPU)301,其可以根据存储在只读存储器(ROM)302中的程序或者从存储部分303加载到随机访问存储器(RAM)303中的程序而执行各种适当的动作和处理。在RAM 303中,还存储有系统300操作所需的各种程序和数据。CPU 301、ROM 302以及RAM 303通过总线304彼此相连。输入/输出(I/O)接口305也连接至总线304。

[0113] 以下部件连接至I/O接口305:包括键盘、鼠标等的输入部分306;包括诸如阴极射线管(CRT)、液晶显示器(LCD)等以及扬声器等的输出部分307;包括硬盘等的存储部分308;以及包括诸如LAN卡、调制解调器等的网络接口卡的通信部分309。通信部分309经由诸如因特网的网络执行通信处理。驱动器310也根据需要连接至I/O接口305。可拆卸介质311,诸如磁盘、光盘、磁光盘、半导体存储器等等,根据需要安装在驱动器310上,以便于从其上读出的计算机程序根据需要被安装入存储部分308。

[0114] 特别地,根据本申请的实施例,上文参考流程图描述的过程可以被实现为计算机软件程序。例如,本申请的实施例包括一种计算机程序产品,其包括承载在机器可读介质上的计算机程序,该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中,该计算机程序可以通过通信部分303从网络上被下载和安装,和/或从可拆卸介质311被安装。在该计算机程序被中央处理单元(CPU) 301执行时,执行本申请的系统中限定的上述功能。

[0115] 需要说明的是,本申请所示的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本申请中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本申请中,计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式,包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质,该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括但不限于:无线、电线、光缆、RF等等,或者上述的任意合适的组合。

[0116] 附图中的流程图和框图,图示了按照本申请各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分,前述模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个接连地表示的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0117] 描述于本申请实施例中所涉及到的单元或模块可以通过软件的方式实现,也可以通过硬件的方式来实现。所描述的单元或模块也可以设置在处理器中,例如,可以描述为:一种处理器,包括:获取模块、模型构建模块及网络构建模块。其中,这些单元或模块的名称在某种情况下并不构成对该单元或模块本身的限定,例如,获取模块还可以被描述为“用于获取待处理数据”。

[0118] 作为另一方面,本申请还提供了一种计算机可读存储介质,该计算机可读存储介质可以是上述实施例中描述的设备中所包含的;也可以是单独存在,而未装配入该设备中的。上述计算机可读存储介质存储有一个或者多个程序,当上述前述程序被一个或者一个以上的处理器用来执行描述于本申请的基于动态本体的知识图谱分析方法:

[0119] 获取待处理数据；

[0120] 对所述待处理数据进行本体化映射处理,构建动态本体模型,所述动态本体模型用于将所述待处理数据的原始存储格式转化为知识库的对象存储格式；

[0121] 基于所述动态本体模型,构建知识图谱网络,所述知识图谱网络用于对所述待处理数据进行分析。

[0122] 综上所述,本申请实施例中提供的基于动态本体的知识图谱分析、应用方法、平台、设备及存储介质,通过获取待处理数据,并对待处理数据进行本体化映射处理,构建动态本体模型,该动态本体模型用于将待处理数据的原始存储方式转化为知识库的对象存储方式,并基于动态本体模型,构建知识图谱网络,以使得通过知识图谱网络对待处理数据进行处理分析。该技术方案能够对海量的多源异构数据进行融合、关联、标注以及知识化处理,构建形成大规模领域关系图谱网络,从而提高了数据深化应用能力,实现了数据的接入获取、存储、计算以及大规模的数据挖掘分析,进而能够实现从数据到知识到智能的升级转变,提高了对数据价值的挖掘效率,能够有效支撑各领域信息化向智能化应用的发展。

[0123] 以上描述仅为本申请的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解,本申请中所涉及的发明范围,并不限于上述技术特征的特定组合而成的技术方案,同时也应涵盖在不脱离所述发明构思的情况下,由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本申请中公开的(但不限于)具有类似功能的技术特征进行互相替换而形成的技术方案。

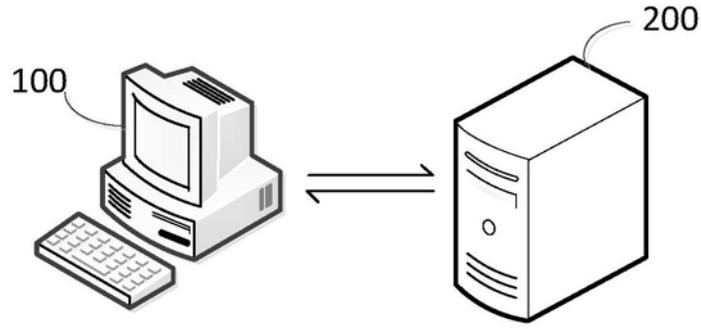


图1

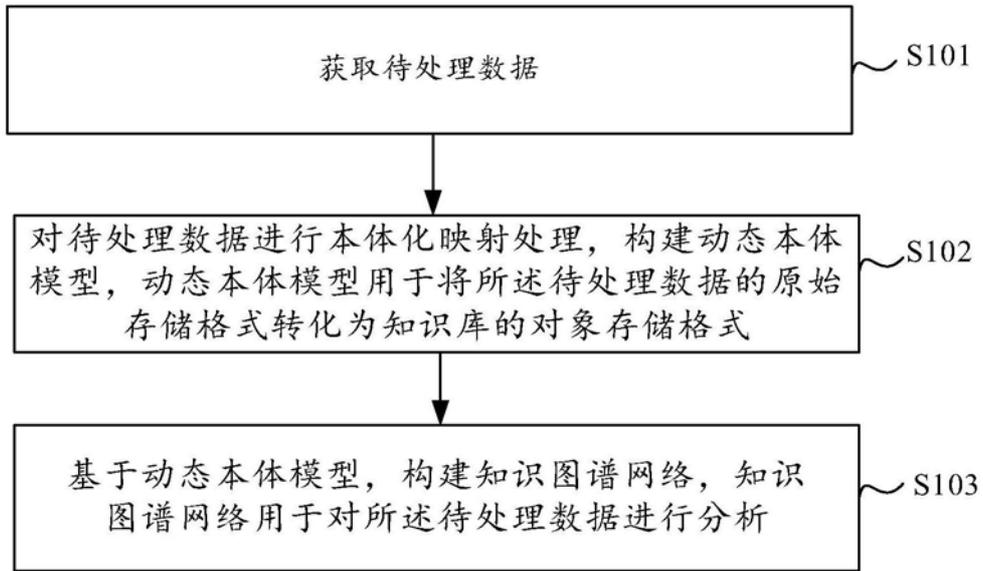


图2

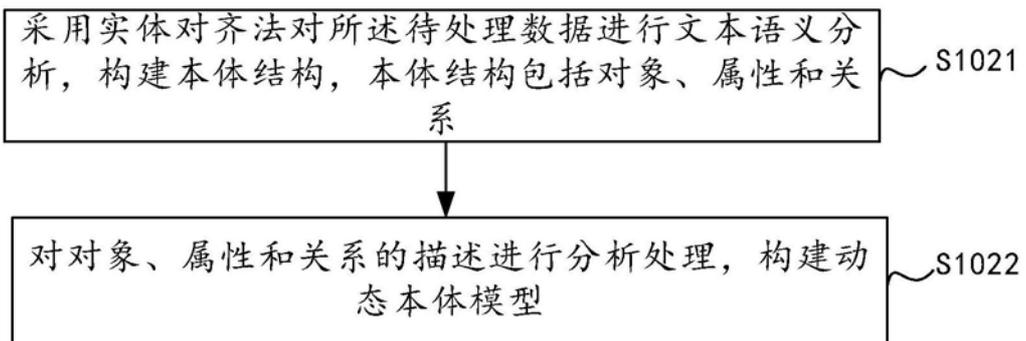


图3

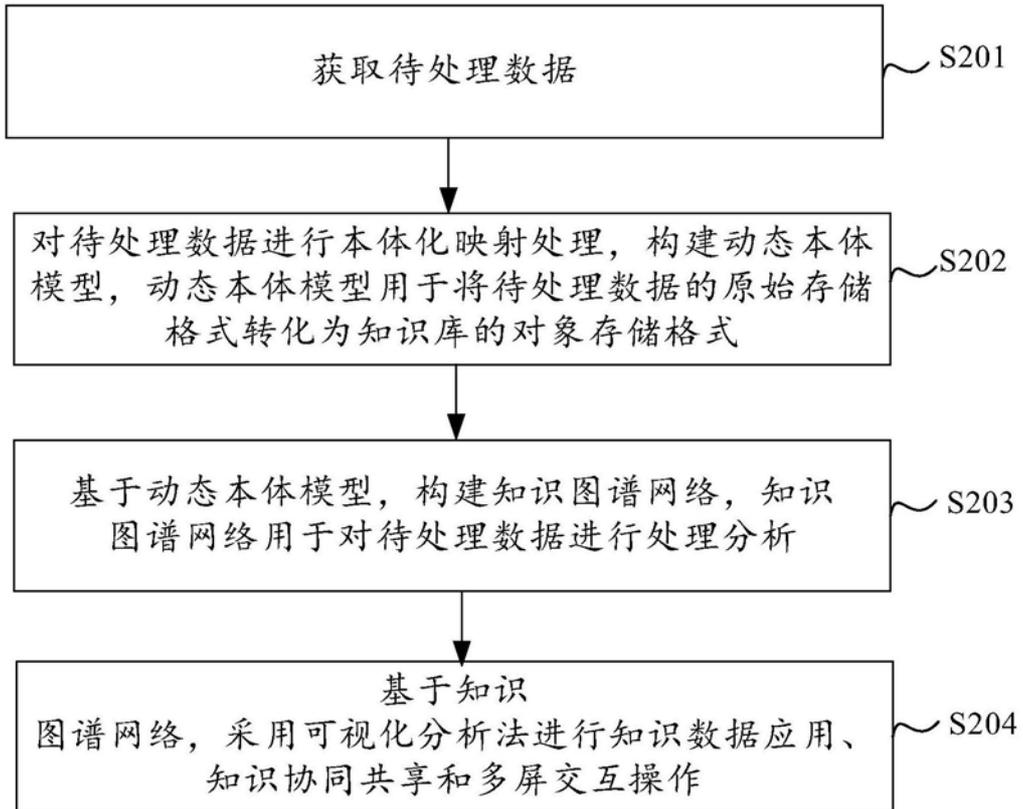


图4

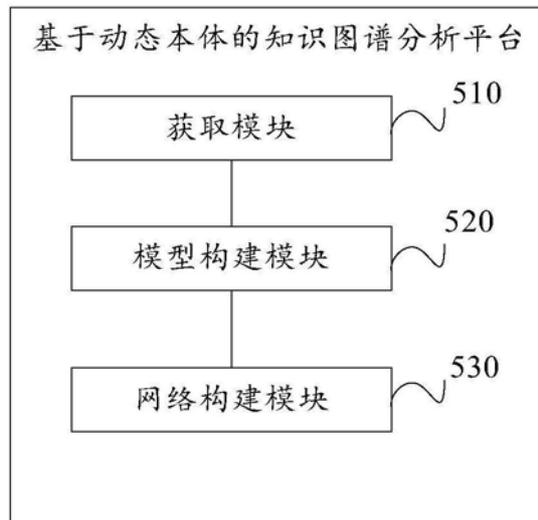


图5

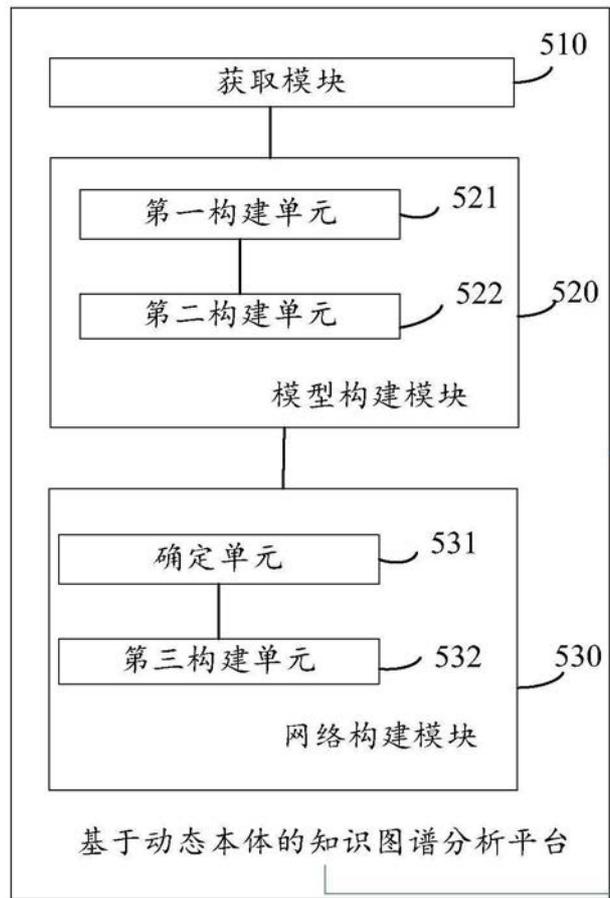


图6

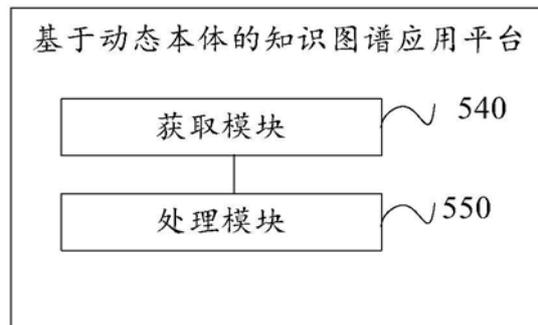


图7

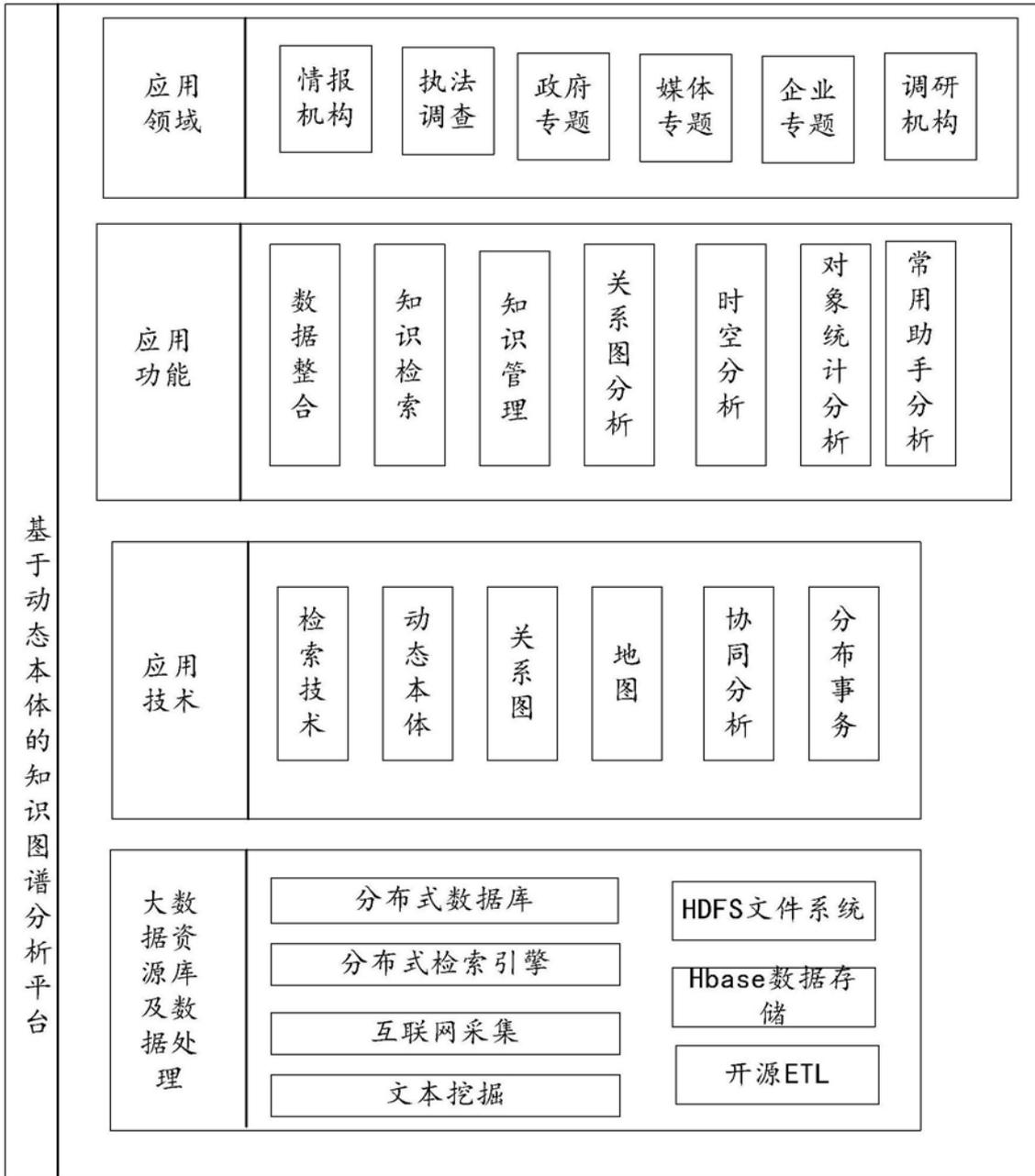


图8

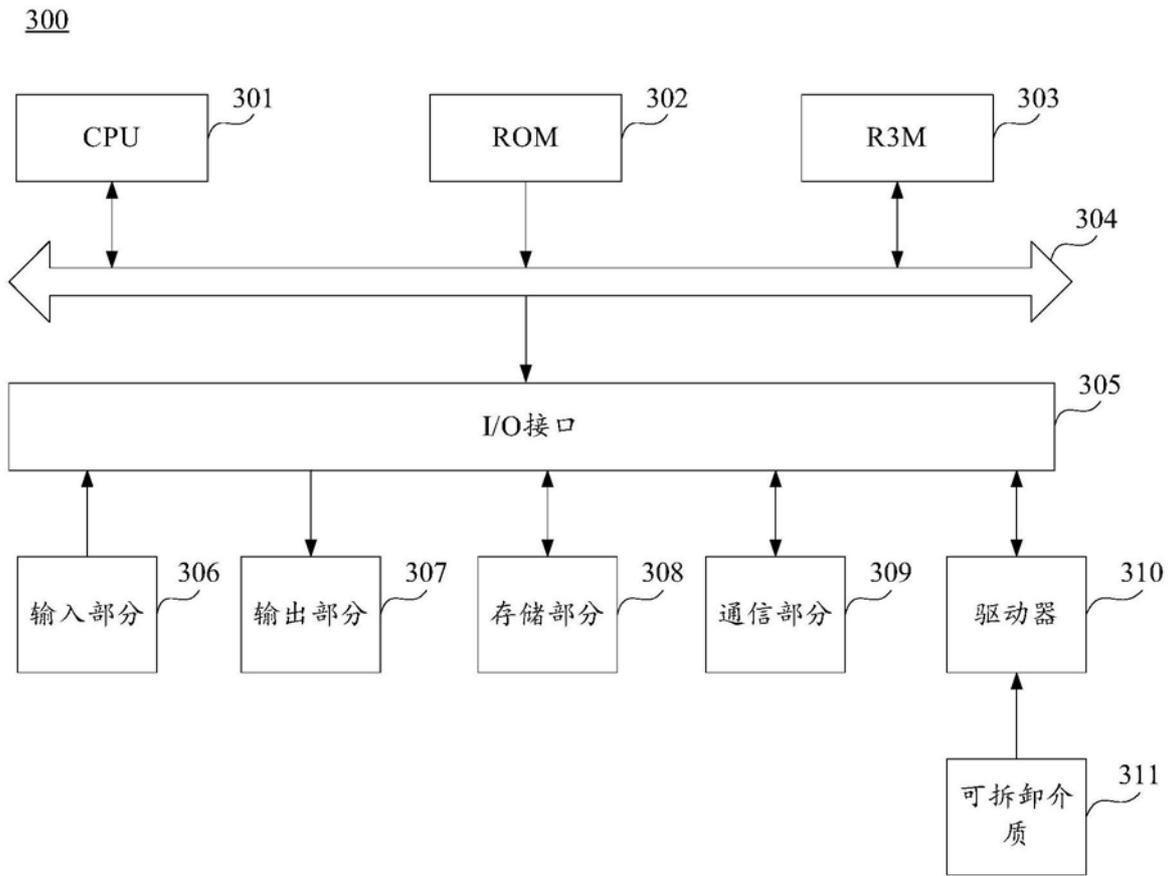


图9