



(12) 发明专利申请

(10) 申请公布号 CN 114443854 A

(43) 申请公布日 2022. 05. 06

(21) 申请号 202111646837.X

(22) 申请日 2021.12.30

(71) 申请人 深圳晶泰科技有限公司

地址 518000 广东省深圳市福田区福保街
道福保社区红柳道2号顺丰工业厂房3
层

(72) 发明人 谈樑 李柄坤 朱和胜 康晓琦
刘阳

(74) 专利代理机构 深圳市威世博知识产权代理
事务所(普通合伙) 44280

专利代理师 刘桂兰

(51) Int. Cl.

G06F 16/36 (2019.01)

权利要求书2页 说明书9页 附图5页

(54) 发明名称

多源异构数据的处理方法、装置、计算机设备
及存储介质

(57) 摘要

本发明涉及知识图谱领域技术领域,特别是
涉及多源异构数据的处理方法、装置、计算机设备
及存储介质。方法包括从可信源系统中获取业务
数据;其中,所述可信源系统包括各个业务域的
终端所使用的业务系统;根据所述业务数据构建
知识图谱;其中,所述知识图谱能够指示不同
业务域的业务数据与业务域指标的关系;根据所
述知识图谱,获取待融合的业务域的资源实体及
所述资源实体的关系,以构建业务域知识图谱
库;基于所述业务域知识图谱库执行跨业务域及
跨资源实体的业务域信息同步、融合及共享操
作。本发明能够对于内外部不同组织、跨业务领
域的信息进行高效提取,以及快速地迭代式地整
合为业务知识。



1. 一种多源异构数据的处理方法,其特征在于,包括:

从可信源系统中获取业务数据;其中,所述可信源系统包括各个业务域的终端所使用的业务系统;

根据所述业务数据构建知识图谱;其中,所述知识图谱能够指示不同业务域的业务数据与业务域指标的关系;

根据所述知识图谱,获取待融合的业务域的资源实体及所述资源实体的关系,以构建业务域知识图谱库;

基于所述业务域知识图谱库执行跨业务域及跨资源实体的业务域信息同步、融合及共享操作。

2. 根据权利要求1所述的多源异构数据的处理方法,其特征在于,在所述基于所述业务域知识图谱库执行跨业务域及跨资源实体的业务域信息同步、融合及共享操作之后,还包括:

根据预先配置的各个业务域对应的业务域指标以及可信源指标,对各个业务域的业务数据质量进行校验;其中,

所述业务域指标和所述可信源指标均被配置为由用户定义的对不同业务域的业务数据进行考量的指标,所述可信源指标被用于作为所述业务域指标比对的参考标准。

3. 根据权利要求1所述的多源异构数据的处理方法,其特征在于,所述根据所述业务数据构建知识图谱,包括:

对所述业务数据进行元数据提取,以建立标准化图数据;其中,所述标准化图数据包括元数据实体及实体关系;

根据所述标准化图数据构建业务域知识图谱。

4. 根据权利要求3所述的多源异构数据的处理方法,其特征在于,所述对所述业务数据进行元数据提取,以建立标准化图数据,包括:

在所述业务数据为既有业务数据时,根据业务域指标模型,对所述既有业务数据进行元数据提取,以建立标准化图数据;

其中,所述业务域指标模型被配置有关联各个业务域的业务域指标,每个所述业务域指标均被赋予相应的权重。

5. 根据权利要求4所述的多源异构数据的处理方法,其特征在于,所述对所述业务数据进行元数据提取,以建立标准化图数据,还包括:

在所述业务数据为增量业务数据时,若所述增量业务数据属于全新业务域,则对所述增量业务数据进行元数据提取,以建立新的标准化图数据;

若所述增量业务数据属于既有业务域,则对所述增量业务数据进行元数据提取;并

根据所述增量业务数据对应的业务域指标以及可信源指标,对所述元数据进行校验,根据校验结果以建立新的标准化图数据或增量合并既有的标准化图数据。

6. 根据权利要求3-5任一项所述的多源异构数据的处理方法,其特征在于,在所述对所述业务数据进行元数据提取之后,还包括:

对所述元数据进行预设数据格式转换以及持久化。

7. 根据权利要求1所述的多源异构数据的处理方法,其特征在于,所述资源实体包括业务系统软件信息、嵌入式软件信息,以及硬件设备信息。

8. 一种多源异构数据的处理装置,其特征在于,包括:

获取模块,用于从可信源系统中获取业务数据;其中,所述可信源系统包括各个业务域的终端所使用的业务系统;

第一构建模块,用于根据所述业务数据构建知识图谱;其中,所述知识图谱能够指示不同业务域的业务数据与业务域指标的关系;

第二构建模块,用于根据所述知识图谱,获取待融合的业务域的资源实体及所述资源实体的关系,以构建业务域知识图谱库;

融合模块,用于基于所述业务域知识图谱库执行跨业务域及跨资源实体的业务域信息同步、融合及共享操作。

9. 根据权利要求8所述的多源异构数据的处理装置,其特征在于,还包括:

评价模块,用于根据预先配置的各个业务域对应的业务域指标以及可信源指标,对各个业务域的业务数据质量进行校验;其中,

所述业务域指标和所述可信源指标均被配置为由用户定义的对不同业务域的业务数据进行考量的指标,所述可信源指标被用于作为所述业务域指标比对的参考标准。

10. 一种计算机设备,其特征在于,包括:处理器、存储器以及存储在所述存储器上的计算机程序,所述处理器耦合所述存储器,所述处理器在工作时执行所述计算机程序以实现如权利要求1-7中任一项所述的多源异构数据的处理方法。

11. 一种计算机可读存储介质,其特征在于,所述计算机存储介质存储有计算机指令,当所述计算机指令被计算机执行时,使得所述计算机执行权利要求1-7中任一项所述的多源异构数据的处理方法的指令。

多源异构数据的处理方法、装置、计算机设备及存储介质

技术领域

[0001] 本发明涉及知识图谱领域技术领域,特别是涉及多源异构数据的处理方法、装置、计算机设备及存储介质。

背景技术

[0002] 随着AI技术的快速发展和应用,科技型企业的技术路线变革速度加快,具备优势技术的科技公司通过单业务线单应用场景的支点,向产业链上下游快速产生了虹吸效应。在此背景下,上下游知识信息聚集爆炸,IT产品线需要在短期内快速扩充,IT技术的应用场景也会快速扩大,这对于不同领域的知识获取、融合及分析,提出了高效、辅助决策等需求。

[0003] 伴随业务线的快速扩张兼并而来的,往往是各发展阶段各内外部组织的信息系统整合和信息共享。在复杂产业链的业务流程场景下,传统ERP是最成熟的实践方案。但ERP往往需要大规模的采购、部署、培训以及高成本的用户习惯、管理体系的迁移成本。在一个快速变革的业务流程的场景下,让既有系统的用户快速融入ERP,并在业务线合并后尽快发挥价值,具有很大的难度。SaaS虽然能够起到一定的去ERP效果,提升业务的灵活度,但SaaS同样与组织、用户习惯进行了深度绑定,并且不同的SaaS厂商,让各自的数据在同一企业内变成了一座座孤岛,进而产生了同样的高迁移成本问题。

[0004] 在对现有技术的长期研究及实践中,本发明的发明人发现,现有技术下的研究和应用,如ERP、SaaS或其他,鲜有关注在企业快速发展的场景下企业对于内外部组织、跨业务领域的增量信息的如何高效提取,以及快速地迭代式地整合为业务知识的问题。致使现有技术对于多元异构企业信息的管理中跨领域数据的提取、整理、融合并转化为运营决策的参考指标缺乏有效且高效的手段。

发明内容

[0005] 基于现有技术中存在的问题和缺点,本发明提供一种多源异构数据的处理方法、装置、计算机设备及存储介质,能够对于内外部不同组织、跨业务领域的信息进行高效提取,以及快速地迭代式地整合为业务知识。

[0006] 本申请一个实施例提供一种多源异构数据的处理方法,包括:

[0007] 从可信源系统中获取业务数据;其中,所述可信源系统包括各个业务域的终端所使用的业务系统;

[0008] 根据所述业务数据构建知识图谱;其中,所述知识图谱能够指示不同业务域的业务数据与业务域指标的关系;

[0009] 根据所述知识图谱,获取待融合的业务域的资源实体及所述资源实体的关系,以构建业务域知识图谱库;

[0010] 基于所述业务域知识图谱库执行跨业务域及跨资源实体的业务域信息同步、融合及共享操作。

[0011] 可选的,在所述基于所述业务域知识图谱库执行跨业务域及跨资源实体的业务域

信息同步、融合及共享操作之后,还包括:

[0012] 根据预先配置的各个业务域对应的业务域指标以及可信源指标,对各个业务域的业务数据质量进行校验;其中,

[0013] 所述业务域指标和所述可信源指标均被配置为由用户定义的对不同业务域的业务数据进行考量的指标,所述可信源指标被用于作为所述业务域指标比对的参考标准。

[0014] 可选的,所述根据所述业务数据构建知识图谱,包括:

[0015] 对所述业务数据进行元数据提取,以建立标准化图数据;其中,所述标准化图数据包括元数据实体及实体关系;

[0016] 根据所述标准化图数据构建业务域知识图谱。

[0017] 可选的,所述对所述业务数据进行元数据提取,以建立标准化图数据,包括:

[0018] 在所述业务数据为既有业务数据时,根据业务域指标模型,对所述既有业务数据进行元数据提取,以建立标准化图数据;

[0019] 其中,所述业务域指标模型被配置有关联各个业务域的业务域指标,每个所述业务域指标均被赋予相应的权重。

[0020] 可选的,所述对所述业务数据进行元数据提取,以建立标准化图数据,还包括:

[0021] 在所述业务数据为增量业务数据时,若所述增量业务数据属于全新业务域,则对所述增量业务数据进行元数据提取,以建立新的标准化图数据;

[0022] 若所述增量业务数据属于既有业务域,则对所述增量业务数据进行元数据提取;并

[0023] 根据所述增量业务数据对应的业务域指标以及可信源指标,对所述元数据进行校验,根据校验结果以建立新的标准化图数据或增量合并既有的标准化图数据。

[0024] 可选的,在所述对所述业务数据进行元数据提取之后,还包括:

[0025] 对所述元数据进行预设数据格式转换以及持久化。

[0026] 可选的,所述资源实体包括业务系统软件信息、嵌入式软件信息,以及硬件设备信息。

[0027] 基于同一发明构思,本申请一个实施例还提供一种多源异构数据的处理装置,包括:

[0028] 获取模块,用于从可信源系统中获取业务数据;其中,所述可信源系统包括各个业务域的终端所使用的业务系统;

[0029] 第一构建模块,用于根据所述业务数据构建知识图谱;其中,所述知识图谱能够指示不同业务域的业务数据与业务域指标的关系;

[0030] 第二构建模块,用于根据所述知识图谱,获取待融合的业务域的资源实体及所述资源实体的关系,以构建业务域知识图谱库;

[0031] 融合模块,用于基于所述业务域知识图谱库执行跨业务域及跨资源实体的业务域信息同步、融合及共享操作。

[0032] 可选的,所述的多源异构数据的处理装置,还包括:

[0033] 评价模块,用于根据预先配置的各个业务域对应的业务域指标以及可信源指标,对各个业务域的业务数据质量进行校验;其中,

[0034] 所述业务域指标和所述可信源指标均被配置为由用户定义的对不同业务域的业

务数据进行考量的指标,所述可信源指标被用于作为所述业务域指标比对的参考标准。

[0035] 基于同一发明构思,本申请一个实施例还提供一种计算机设备,包括:处理器、存储器以及存储在所述存储器上的计算机程序,所述处理器耦合所述存储器,所述处理器在工作时执行所述计算机程序以实现如上述的多源异构数据的处理方法。

[0036] 基于同一发明构思,本申请一个实施例还提供一种计算机可读存储介质,所述计算机存储介质存储有计算机指令,当所述计算机指令被计算机执行时,使得所述计算机执行上述的多源异构数据的处理方法的指令。

[0037] 上述技术方案中的一个技术方案具有如下优点和有益效果:

[0038] 本申请各实施例,通过知识抽取技术构建与可信源系统中获取业务数据对应的知识图谱。进一步地,根据所述知识图谱,获取待融合的业务域的资源实体及所述资源实体的关系,以构建业务域知识图谱库。基于所述业务域知识图谱库执行跨业务域及跨资源实体的业务域信息同步、融合及共享操作。基于此,本申请能够在企业的业务流程快速迭代和整合过程中,针对跨领域数据的高效提取、整理、融合并转化为运营决策的参考指标。

附图说明

[0039] 本申请将结合附图对实施方式进行了说明。本申请的附图仅用于描述实施例,以展示为目的。在不偏离本申请原理的条件下,本领域技术人员能够轻松地通过以下描述根据所述步骤做出其他实施例。

[0040] 图1为本申请一个实施例中多源异构数据的管理系统的结构示意图;

[0041] 图2为本申请一个实施例中多源异构数据的处理方法的流程示意图;

[0042] 图3为本申请一个实施例中多源异构数据的处理方法的流程示意图;

[0043] 图4为本申请一个实施例中核心价值流程图谱示意图;

[0044] 图5为本申请一个实施例中增量业务域的图谱示意图;

[0045] 图6为本申请一个实施例中基于业务领域的知识图谱构建过程示意图;

[0046] 图7为本申请一个实施例中多源异构数据的处理方法的流程示意图;

[0047] 图8为本申请一个实施例中基于可信源指标配置单元的图谱原理示意图;

[0048] 图9为本申请一个实施例中多源异构数据的处理装置的结构示意图;

[0049] 图10为本申请一个实施例中多源异构数据的处理装置的结构示意图。

具体实施方式

[0050] 下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚、完整地描述。可以理解的是,此处所描述的具体实施例仅用于解释本申请,而非对本申请的限定。另外还需要说明的是,为了便于描述,附图中仅示出了与本申请相关的部分而非全部结构。基于本申请中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本申请保护的范围。

[0051] 本申请中的术语“第一”、“第二”等是用于区别不同对象,而不是用于描述特定顺序。此外,术语“包括”和“具有”以及它们任何变形,意图在于覆盖不排他的包含。例如包含了一系列步骤或单元的过程、方法、系统、产品或设备没有限定于已列出的步骤或单元,而是可选地还包括没有列出的步骤或单元,或可选地还包括对于这些过程、方法、产品或设备

固有的其它步骤或单元。

[0052] 在本文中提及“实施例”意味着,结合实施例描述的特定特征、结构或特性可以包含在本申请的至少一个实施例中。在说明书中的各个位置出现该短语并不一定均是指相同的实施例,也不是与其它实施例互斥的独立的或备选的实施例。本领域技术人员显式地和隐式地理解的是,本文所描述的实施例可以与其它实施例相结合。

[0053] 图是数据结构和算法学中最强大的框架之一,它是一种由顶点和边所组成的抽象网络。在特定场景中,通过对顶点和边进行合理的定义描述,可以构建出客观世界里各类抽象实体之间关系的语义网络。

[0054] 知识图谱是基于图论的最广泛的应用,常规的知识图谱在逻辑结构上分为模式层和数据层,数据层由一系列(实体、关系、实体)三元组来表达事实。模式层则定义了对于事实的描述规则。一套知识图谱的建立,一般通过知识表达,知识抽取,知识融合来完成。在多源异构的企业信息管理领域,知识图谱和图论的研究主要关注点是企业关系,即通过公开信息的爬取整理,抽取出企业间的关联信息,为尽调和监管提供分析参考。而在企业内部管理领域,研究更多聚焦于知识管理本身,将成熟企业流程化、文化等知识进行图谱构建,进而提高数据查询和流程执行的效率。

[0055] 然而,过往的研究和应用鲜有关注在企业快速发展的场景下,企业对于内外部组织、跨业务领域的增量信息的提取与学习,并且快速和迭代式地整合为业务知识的效率的诉求远远领先于企业信息整合,知识图谱在此领域可以发挥更大的价值。

[0056] 基于此,在本发明的一个实施例中,提供一种多源异构数据的管理系统。

[0057] 如图1所示,多源异构数据的管理系统,包括中心化的业务域价值指标设计和应用模块、分布式多源知识提取架构和多源业务域图谱评价模块,它们与整体IT服务中的提供元数据的现存系统及潜在引入的新系统形成了一整套的分析体系架构。

[0058] 业务域价值指标设计和应用模块,包括业务域指标和模型设计单元、数据描述规则设计单元、业务域知识图谱可视化单元、业务域指标变更日志单元和智能决策报表引擎。

[0059] 其中,业务域指标和模型设计单元为用户提供了可视化的指标设计工具。用户可以定义、拖拽以及关联各类指标并形成图结构的业务域指标模型。设计完成的业务域指标模型可以指导智能决策报表引擎,对分布式多源知识提取架构中存储的数据进行提取和分析,形成可读报表。

[0060] 数据描述规则设计单元用以统一指标语言,方便用户理解和系统解析。

[0061] 业务域知识图谱可视化单元宏观展示了不同业务域的数据与指标关联关系,方便用户快速学习业务域知识。

[0062] 业务域指标变更日志单元记录了业务域指标模型变更,方便用户追溯过去的版本。

[0063] 分布式多源知识提取架构,包括可信源系统数据标注与提取单元、可信源系统管理单元、数据同步和归集服务和分布式数据存储管理。

[0064] 其中,可信源系统数据标注和提取单元的主要作用是对来自可信源系统的数据进行提取和处理,处理功能包括将元数据进行抽取并统一转换成数据描述规则设计中定义的数据格式。

[0065] 可信源系统管理对可以接入分布式多源知识提取架构的元数据来源系统进行管

理,确定哪一些系统可以被纳入架构,是否有数据同步的权限。

[0066] 数据同步和归集服务单元定期将分布式的数据进行备份和清洗,以保证微服务的稳定性和可靠性。

[0067] 分布式数据存储管理单元对统一格式的数据进行了持久化,以保证上层应用对数据的分析需求。

[0068] 多元业务域评价模块包括可信源指标配置单元、业务域指标比对和评价单元。

[0069] 其中,可信源指标配置单元允许用户对不同领域的系统进行指标设计,并作为业务域指标比对和评价单元的参考标准。

[0070] 业务域指标比对和评价单元在新图谱创建的时候,会与可信源指标配置单元中存储的参考图谱进行比对,对用户输出评价。

[0071] 可信源系统,作为各个业务域的终端所使用的业务系统,包括但不限于项目管理系统、销售管理系统,实验室LIMS、供应链管理系统、人力资源系统,以及其他业务系统。可以理解的是,所述可信源系统支撑起了终端用户的日常工作,并且存储了大量的孤岛数据。

[0072] 在一个实施例中,由于整合业务线的不同IT系统的数据源不仅存在数据结构的差异,还存在地域、网络配置的区别。因此,分布式多源知识提取架构,为中心化管理、分布式服务部署的模式,需要针对不同的数据源系统进行相应的配置管理。

[0073] 基于此,针对分布式多源知识提取架构,对可信源系统设计了分布式数据存储和管理功能,以分治相关的系统配置和追溯查询元数据。

[0074] 可信源系统的定义是整体IT系统的权限认证并可以进行数据对接的系统。它通过多源知识提取架构的中心化管理后台进行配置和赋权,并通过匹配不同环境中部署的多源知识提取服务实现数据提取、标注和清洗等工作程序。

[0075] 分布式的多源知识提取服务提供元数据的查询、检索等功能,提供接口给相关应用层的服务对数据进行数据的分析或者追溯。

[0076] 上述实施例实施方式充分结合知识图谱技术,运用在快速发展过程中积累的业务域成功经营指标和经验体系,能够通过知识抽取技术构建业务域知识图谱库,将核心经营的业务域实体及实体关系导入,执行跨业务域,业务线的数据挖掘,实现多元异构的业务线指标数据的快速整合和可视化分析展示。

[0077] 如图2所示,基于上述实施例,本申请一个实施例提供一种多源异构数据的处理方法,包括步骤S100-步骤S400。

[0078] 步骤S100:从可信源系统中获取业务数据。其中,所述可信源系统包括各个业务域的终端所使用的业务系统。

[0079] 步骤S200:根据所述业务数据构建知识图谱。其中,所述知识图谱能够指示不同业务域的业务数据与业务域指标的关系。

[0080] 如图3所示,步骤S200包括:

[0081] 步骤S211:对所述业务数据进行元数据提取,以建立标准化图数据。其中,所述标准化图数据包括元数据实体及实体关系。

[0082] 可以理解的是,在对所述业务数据进行元数据提取之后,需要对所述元数据进行预设数据格式转换以及持久化。

[0083] 步骤S212:根据所述标准化图数据构建业务域知识图谱。

[0084] 在一个实施例中,在所述业务数据为既有业务数据时,根据业务域指标模型,对所述既有业务数据进行元数据提取,以建立标准化图数据。

[0085] 其中,所述业务域指标模型被配置有关联各个业务域的业务域指标,每个所述业务域指标均被赋予相应的权重。

[0086] 围绕具体的业务线融合方式,基于图1中业务域知识图谱可视化单元,可以快速构建出决策所需的指标层次以及关联业务域的考量指标。例如:

[0087] 业务线整合时构建效率指标模型时,可以通过人工定义或者决策系统导入的方式,在业务域指标和模型设计的人机界面,定义该模型的名称,版本以及关联的各个业务领域的效率相关指标,并赋予不同的权重,进而形成了经营指标-权重-业务域指标的三元组。进一步地,指导分布式多源知识提取架构对业务域元数据进行提取,以及指导元数据所存在的系统的提取模式与规则进行设计。

[0088] 具体到业务域指标中,与分布式多源知识提取架构归集的数据会呈现在可视化的图谱面板中,供分析系统的用户进行拖拽,关联和制定,最后形成一套基于图的业务线经营指标模型,如图4所示。

[0089] 在一个实施例中,在所述业务数据为增量业务数据时,若所述增量业务数据属于全新业务域,则对所述增量业务数据进行元数据提取,以建立新的标准化图数据。若所述增量业务数据属于既有业务域,则对所述增量业务数据进行元数据提取。

[0090] 根据所述增量业务数据对应的业务域指标以及可信源指标,对所述元数据进行校验。

[0091] 根据校验结果以建立新的标准化图数据或增量合并既有的标准化图数据。

[0092] 如图5所示,图5为增量业务域的图谱示意图。当业务线开始整合时,针对增量引入的信息系统,通过多源提取架构对数据进行提取和归集。第一步是构建增量信息系统的知识图谱,若它是一个全新的领域,则需要单独构建新的图谱。若它是一个已有的领域,则可以首先进行数据的提取和比对,形成多元业务域的具体评价,再由多源信息提取架构决定重新构建图谱还是增量合并图谱。

[0093] 如图6所示,基于业务领域的知识图谱构建过程可以举例如下。标准化的图数据构建完成后,基于对应的业务域指标模型,即可形成知识图谱。通过图1中智能决策报表引擎进行数据的清洗和计算(例如,基于业务域指标模型里的具体设计:业务域指标实体,与其关联的指标或者元数据实体,以及三元组的连接关系权重,来对存储的数据进行计算,最后在用户界面上生成报表),即可以得出增量的业务域在当前待整合业务线对整体经营的贡献。用户通过对指标的观测和分析,使用查询工具和图谱追溯工具,即可了解具体哪一条系统指标影响了整体指标或者产生了正向的收益。

[0094] 步骤S300:根据所述知识图谱,获取待融合的业务域的资源实体及所述资源实体的关系,以构建业务域知识图谱库。

[0095] 在一个实施例中,所述资源实体包括业务系统软件信息、嵌入式软件信息,以及硬件设备信息。

[0096] 步骤S400:基于所述业务域知识图谱库执行跨业务域及跨资源实体的业务域信息同步、融合及共享操作。

[0097] 如图7所示,在步骤S400之后,还包括步骤S500。

[0098] 步骤S500:根据预先配置的各个业务域对应的业务域指标以及可信源指标,对各个业务域的业务数据质量进行校验。

[0099] 其中,所述业务域指标和所述可信源指标均被配置为由用户定义的对不同业务域的业务数据进行考量的指标,所述可信源指标被用于作为所述业务域指标比对的参考标准。

[0100] 如图8所示,图8为基于可信源指标配置单元的图谱原理。一个可信源系统内的信息结构往往代表着一个业务领域的成熟方法论,因此可以围绕该可信源系统和其所在的行业构建一套量化评价方法,其同样是基于图数据的知识图谱。在新引入分析系统的增量数据源中,构建整合的业务域指标模型要考察其评价分数,以供决策者判断业务域指标模型的合理性。

[0101] 基于上述实施例,本申请能够在企业的业务流程快速迭代和整合过程中,针对跨领域数据的高效提取、整理、融合并转化为运营决策的参考指标。

[0102] 如图9所示,基于同一发明构思,本申请一个实施例还提供一种多源异构数据的处理装置,包括:

[0103] 获取模块10,用于从可信源系统中获取业务数据。其中,所述可信源系统包括各个业务域的终端所使用的业务系统。

[0104] 第一构建模块20,用于根据所述业务数据构建知识图谱。其中,所述知识图谱能够指示不同业务域的业务数据与业务域指标的关系。

[0105] 所述第一构建模块20,用于对所述业务数据进行元数据提取,以建立标准化图数据,并根据所述标准化图数据构建业务域知识图谱。其中,所述标准化图数据包括元数据实体及实体关系。

[0106] 可以理解的是,在对所述业务数据进行元数据提取之后,需要对所述元数据进行预设数据格式转换以及持久化。

[0107] 在一个实施例中,所述第一构建模块20,还用于在所述业务数据为既有业务数据时,根据业务域指标模型,对所述既有业务数据进行元数据提取,以建立标准化图数据。其中,所述业务域指标模型被配置有关联各个业务域的业务域指标,每个所述业务域指标均被赋予相应的权重。

[0108] 围绕具体的业务线融合方式,基于图1中业务域知识图谱可视化单元,可以快速构建出决策所需的指标层次以及关联业务域的考量指标。例如:

[0109] 业务线整合时构建效率指标模型时,可以通过人工定义或者决策系统导入的方式,在业务域指标和模型设计的人机界面,定义该模型的名称,版本以及关联的各个业务领域的效率相关指标,并赋予不同的权重,进而形成了经营指标-权重-业务域指标的三元组。进一步地,指导分布式多源知识提取架构对业务域元数据进行提取,以及指导元数据所存在的系统的提取模式与规则进行设计。

[0110] 具体到业务域指标中,与分布式多源知识提取架构归集的数据会呈现在可视化的图谱面板中,供分析系统的用户进行拖拽,关联和制定,最后形成一套基于图的业务线经营指标模型,如图4所示。

[0111] 在一个实施例中,所述第一构建模块20,还用于:

[0112] 在所述业务数据为增量业务数据时,若所述增量业务数据属于全新业务域,则对

所述增量业务数据进行元数据提取,以建立新的标准化图数据;

[0113] 若所述增量业务数据属于既有业务域,则对所述增量业务数据进行元数据提取;

[0114] 根据所述增量业务数据对应的业务域指标以及可信源指标,对所述元数据进行校验。

[0115] 根据校验结果以建立新的标准化图数据或增量合并既有的标准化图数据。

[0116] 如图5所示,图5为增量业务域的图谱示意图。当业务线开始整合时,针对增量引入的信息系统,通过多源提取架构对数据进行提取和归集。第一步是构建增量信息系统的知识图谱,若它是一个全新的领域,则需要单独构建新的图谱。若它是一个已有的领域,则可以首先进行数据的提取和比对,形成多元业务域的具体评价,再由多源信息提取架构决定重新构建图谱还是增量合并图谱。

[0117] 如图6所示,基于业务领域的知识图谱构建过程可以举例如下。标准化的图数据构建完成后,基于对应的业务域指标模型,即可形成知识图谱。通过图1中智能决策报表引擎进行数据的清洗和计算(例如,基于业务域指标模型里的具体设计:业务域指标实体,与其关联的指标或者元数据实体,以及三元组的连接关系权重,来对存储的数据进行计算,最后在用户界面上生成报表),即可以得出增量的业务域在当前待整合业务线对整体经营的贡献。用户通过对指标的观测和分析,使用查询工具和图谱追溯工具,即可了解具体哪一条系统指标影响了整体指标或者产生了正向的收益。

[0118] 第二构建模块30,用于根据所述知识图谱,获取待融合的业务域的资源实体及所述资源实体的关系,以构建业务域知识图谱库。

[0119] 在一个实施例中,所述资源实体包括业务系统软件信息、嵌入式软件信息,以及硬件设备信息。

[0120] 融合模块40,用于基于所述业务域知识图谱库执行跨业务域及跨资源实体的业务域信息同步、融合及共享操作。

[0121] 如图10所示,所述的多源异构数据的处理装置,还包括:

[0122] 评价模块50,用于根据预先配置的各个业务域对应的业务域指标以及可信源指标,对各个业务域的业务数据质量进行校验;其中,

[0123] 所述业务域指标和所述可信源指标均被配置为由用户定义的对不同业务域的业务数据进行考量的指标,所述可信源指标被用于作为所述业务域指标比对的参考标准。

[0124] 如图7所示,图7为基于可信源指标配置单元的图谱原理。一个可信源系统内的信息结构往往代表着一个业务领域的成熟方法论,因此可以围绕该可信源系统和其所在的行业构建一套量化评价方法,其同样是基于图数据的知识图谱。在新引入分析系统的增量数据源中,构建整合的业务域指标模型要考察其评价分数,以供决策者判断业务域指标模型的合理性。

[0125] 本申请一个实施例提供一种计算机设备,包括:处理器、存储器以及存储在所述存储器上的计算机程序,所述处理器耦合所述存储器,所述处理器在工作时执行所述计算机程序以实现如上述的多源异构数据的处理方法。

[0126] 本申请一个实施例提供一种计算机可读存储介质,所述计算机存储介质存储有计算机指令,当所述计算机指令被计算机执行时,使得所述计算机执行如上述的多源异构数据的处理方法的指令。

[0127] 在上述实施例中,可以全部或部分地通过软件、硬件、固件或者其任意结合来实现。当使用软件实现时,可以全部或部分地以计算机程序产品的形式实现。所述计算机程序产品包括一个或多个计算机指令。在计算机上加载和执行所述计算机指令时,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个计算机可读存储介质传输,例如,所述计算机指令可以从一个网站站点、计算机、服务器或数据中心通过有线(例如:同轴电缆、光纤、数据用户线(Digital Subscriber Line,DSL))或无线(例如:红外、无线、微波等)方式向另一个网站站点、计算机、服务器或数据中心进行传输。所述计算机可读存储介质可以是计算机能够存取的任何可用介质,或者是包含一个或多个可用介质集成的服务器、数据中心等数据存储设备。所述可用介质可以是磁性介质(例如:软盘、硬盘、磁带)、光介质(例如:数字通用光盘(Digital Versatile Disc,DVD))或半导体介质(例如:固态硬盘(Solid State Disk,SSD))等。

[0128] 以上所述为本申请提供的实施例,并不用以限制本申请,凡在本申请的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本申请的保护范围之内。

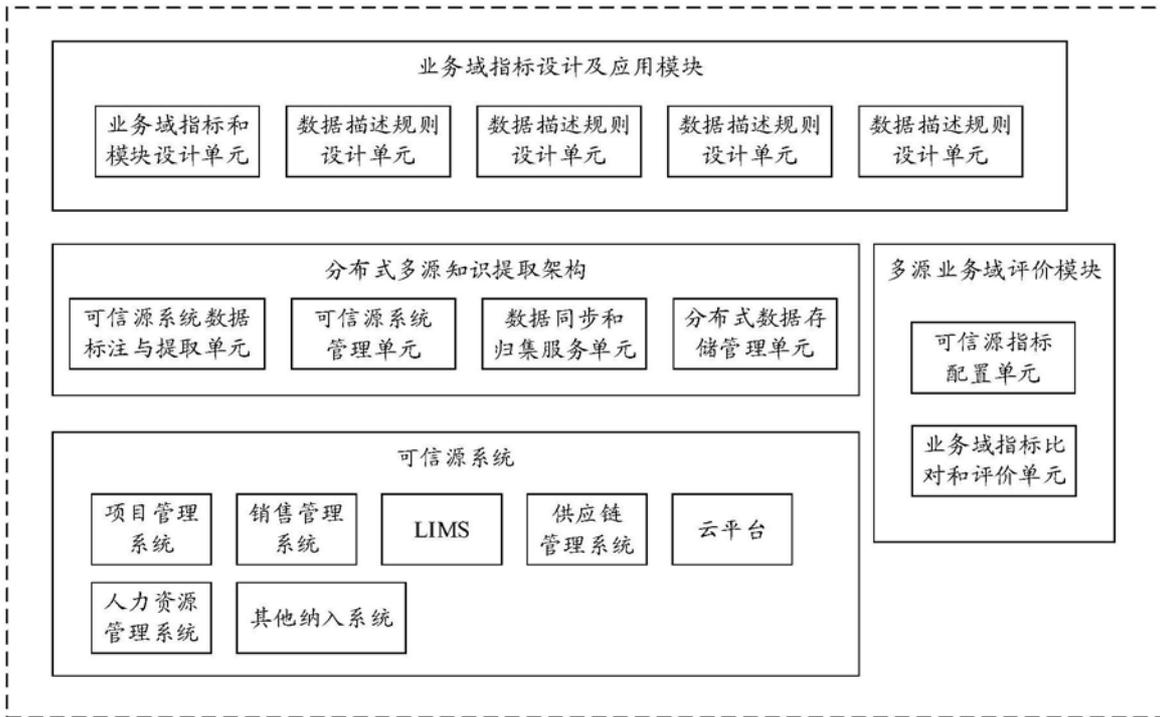


图1

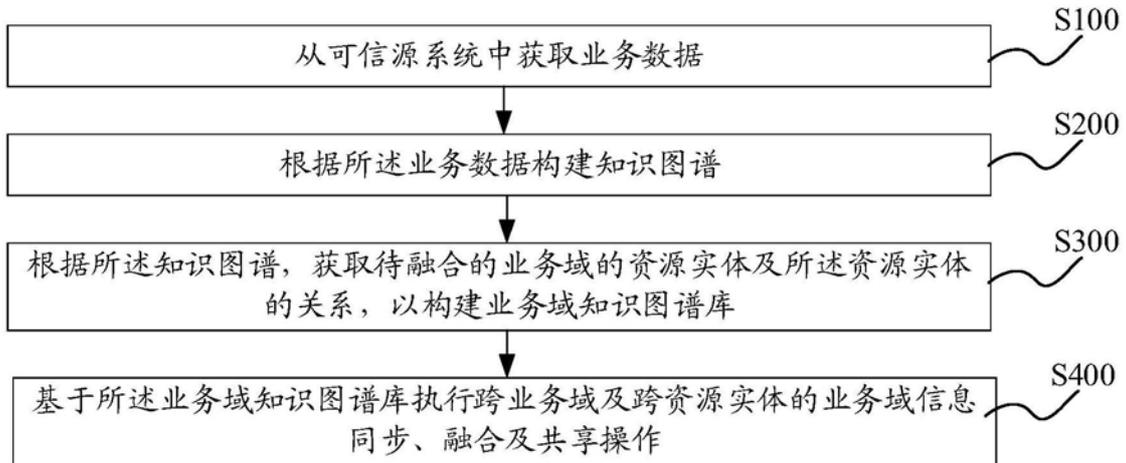


图2

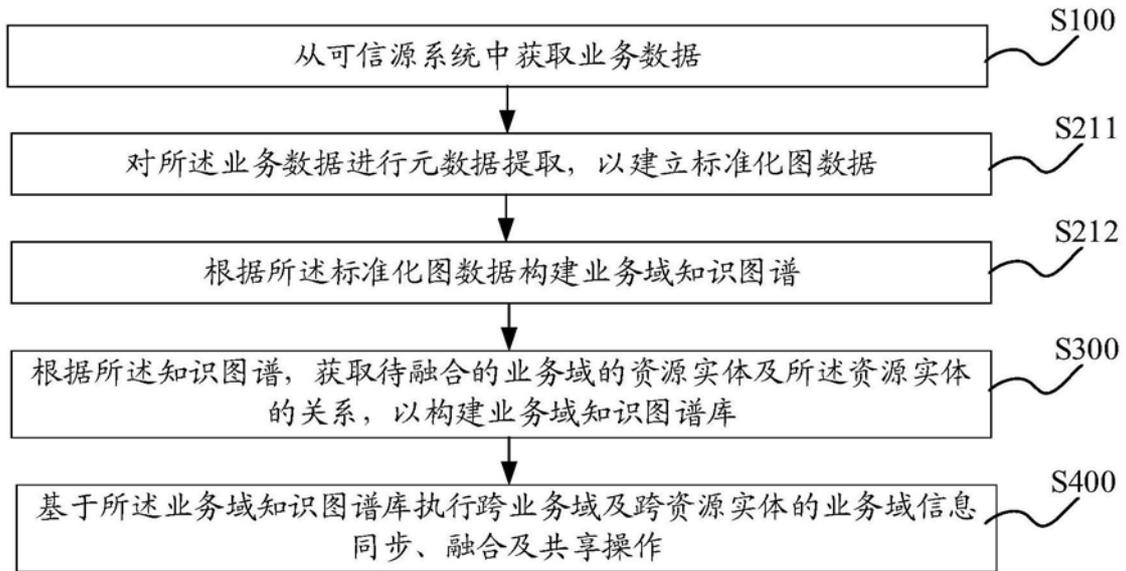


图3

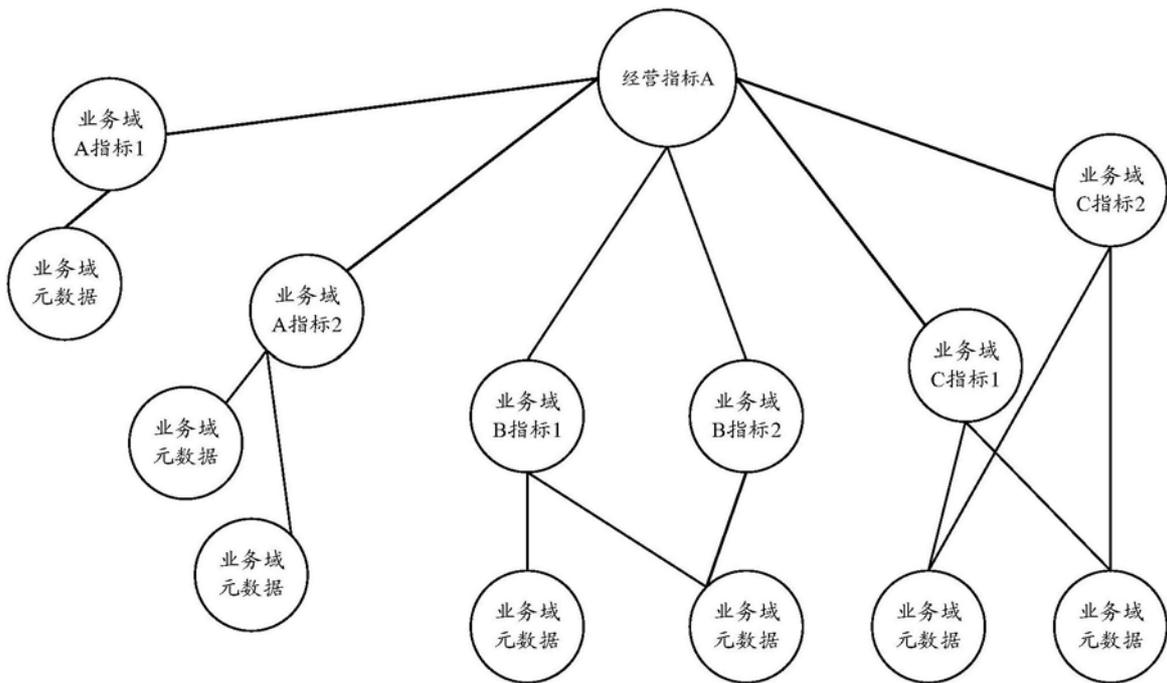


图4

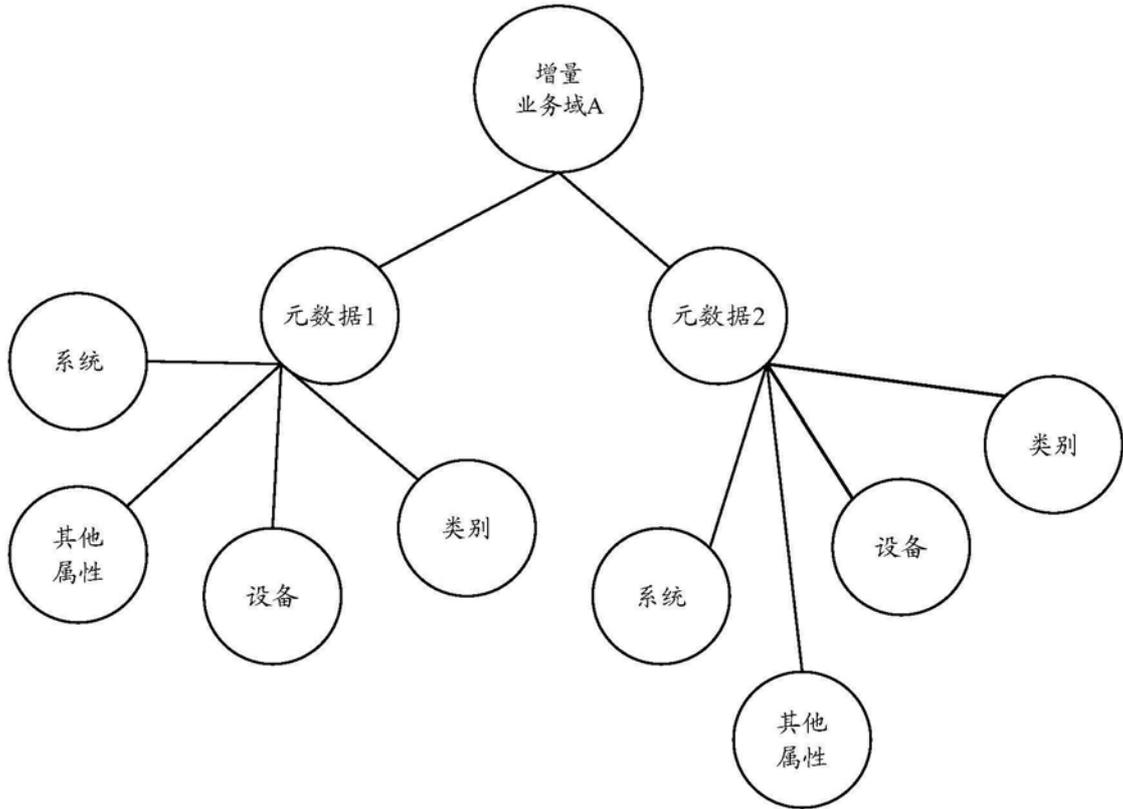


图5

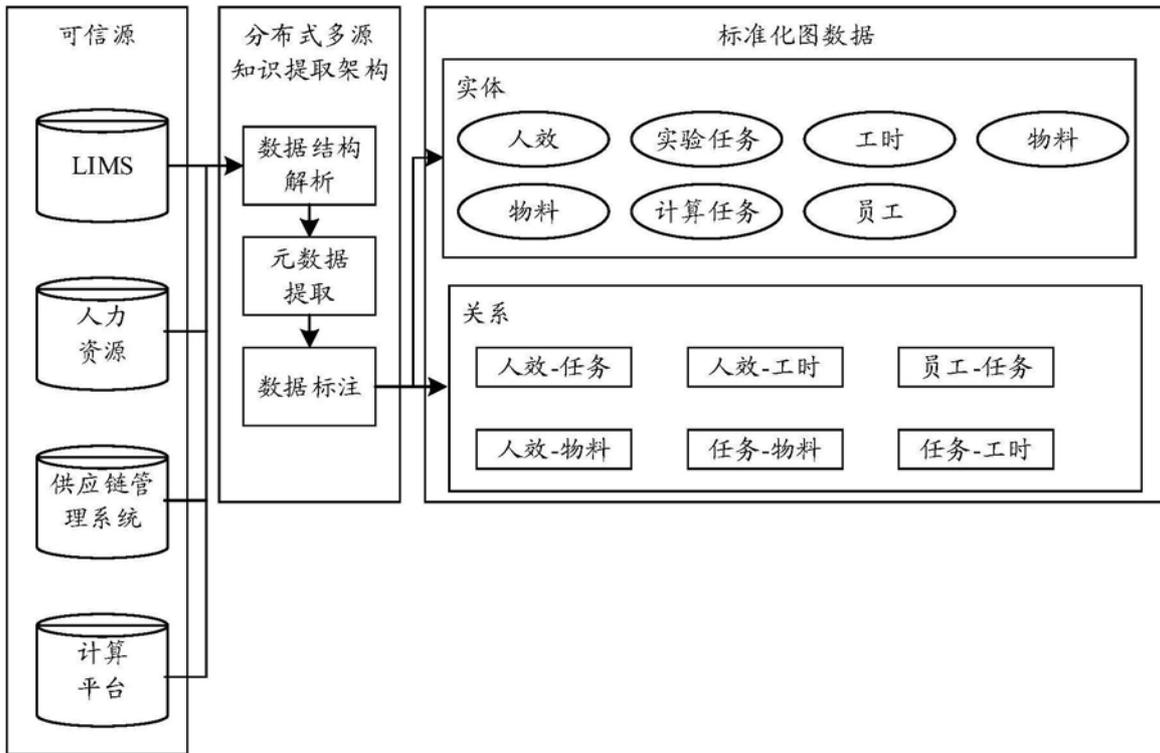


图6

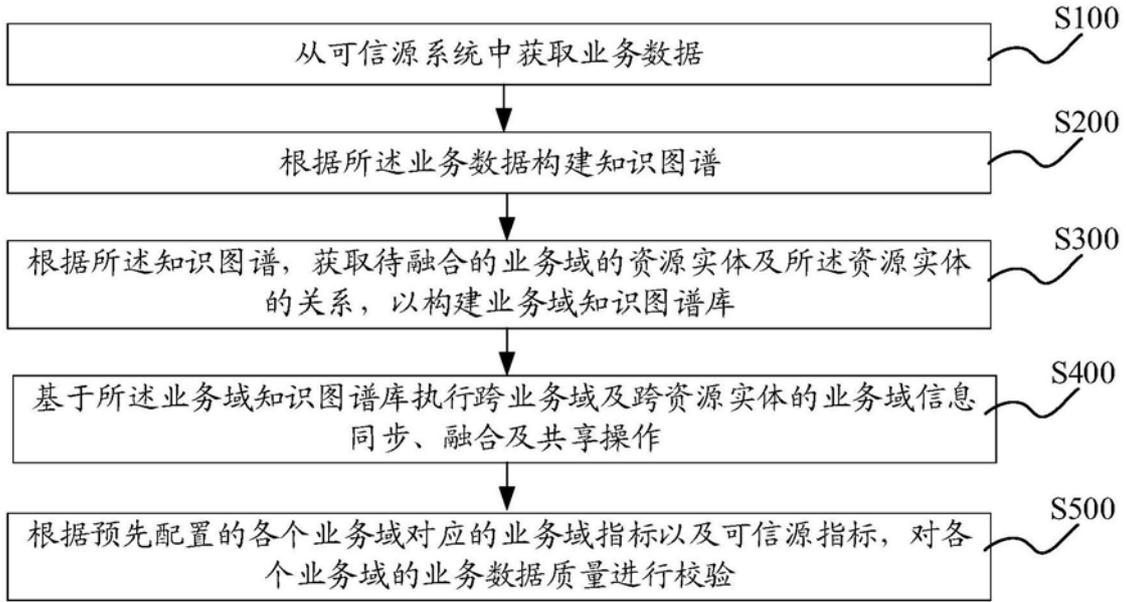


图7

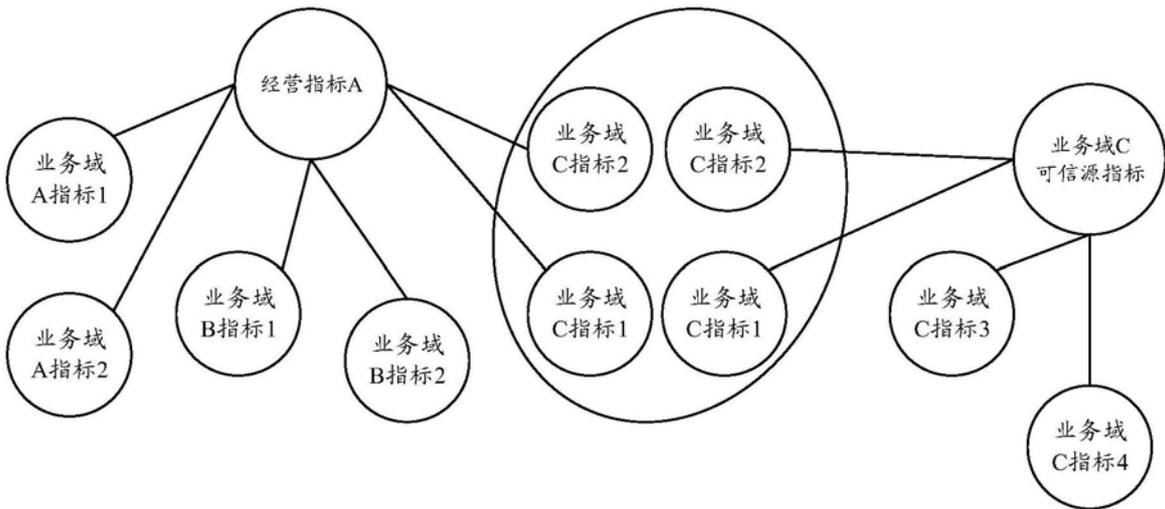


图8

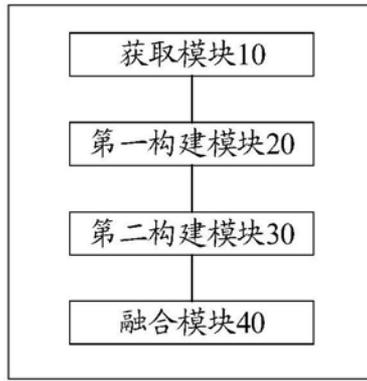


图9

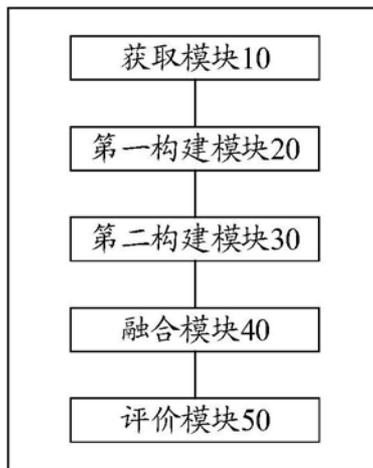


图10