



(12) 发明专利申请

(10) 申请公布号 CN 116028467 A

(43) 申请公布日 2023. 04. 28

(21) 申请号 202211431496.9

(22) 申请日 2022.11.15

(71) 申请人 珠海市新德汇信息技术有限公司
地址 519000 广东省珠海市高新区唐家湾镇科技三路19号1栋

(72) 发明人 杨刚 陈纯冰 陈述 龚波
苏学武 水军 曾志梁

(74) 专利代理机构 广东朗乾律师事务所 44291
专利代理师 侯丽燕 田学东

(51) Int. Cl.

G06F 16/21 (2019.01)

G06F 16/215 (2019.01)

G06F 16/2455 (2019.01)

G06F 16/28 (2019.01)

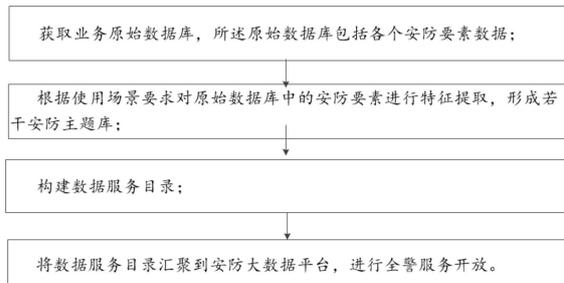
权利要求书2页 说明书12页 附图2页

(54) 发明名称

智能服务大数据建模方法、系统、存储介质及计算机设备

(57) 摘要

本发明提供了智能服务大数据建模方法、系统、存储介质及计算机设备,方法包括:获取业务原始数据库,所述原始数据库包括各个安防要素数据,然后根据使用场景要求对原始数据库中的安防要素进行特征提取,形成若干安防主题库;构建数据服务目录;最后将数据服务目录汇聚到大数据平台,进行安防服务开放。本发明实现了内外部资源的汇聚,提升了安防打防管控和内部管理水平,同时依托大数据建模,提升了安防研判分析能力,拓展了智慧民生,提升了安防管控的民生服务水平。



1. 一种智能服务大数据建模方法,其特征在于,包括:
获取业务原始数据库,所述原始数据库包括各个安防要素数据;
根据使用场景要求对原始数据库中的安防要素进行特征提取,形成若干主题库;
构建数据服务目录;
将数据服务目录汇聚到大数据平台,进行安防服务开放。
2. 如权利要求1所述的智能服务大数据建模方法,其特征在于,获取业务原始数据库的方法包括:
获取各个离散的、多元多态的独立文件并汇聚形成原始数据项;
对所述原始数据项通过大数据平台进行清洗加工,获得所述业务原始数据库。
3. 如权利要求2所述的智能服务大数据建模方法,其特征在于,所述大数据平台通过PL/SQL Developer、TOAD或Navicat Premium进行数据清洗加工,对数据进行清洗加工包括:
对数据进行规则定义,包括对数据进行基础信息的定义、数据映射关系的定义、数据分级分类的定义和数据质量核验规则的定义中的至少一种操作;
和/或根据关键字段进行数据提取;
和/或根据数据定义对数据进行数据过滤、去重、格转、效验中的至少一种操作;
和/或还包括将数据通过分治法、动态规划法、贪心算法、回溯法、分支限界法中的至少一种算法与其他相关业务对应的进行关联回填、关联提取、关联分析;
和/或与大数据平台的数据进行结构化和/或非结构化对比。
4. 如权利要求1所述的智能服务大数据建模方法,其特征在于,所述主题库包括安防目标人员主题库、场所主题库、物品主题库、案件主题库、事件主题库、组织主题库。
5. 如权利要求1所述的智能服务大数据建模方法,其特征在于,构建数据服务目录的方法包括:
根据不同的业务应用需要并按照分层解耦的方法构建数据服务目录;
根据数据服务目录中设置满足业务直接调用集成的不同的数据服务接口,可通过不同的数据服务接口调用不同的业务应用。
6. 如权利要求5所述的智能服务大数据建模方法,其特征在于,所述数据服务接口还设置有访问规则和权限,方便进行数据鉴权,数据鉴权的维度包括内容敏感度、数据来源、数据种类、字段及字段关系分类。
7. 如权利要求1所述的智能服务大数据建模方法,其特征在于,安防服务包括将安防数据按照数据敏感级别分成不同等级,进行分级授权,还包括提供模型分析及数据推送服务,其中,模型分析服务包括根据业务需要,利用分治法、动态规划法、贪心算法、回溯法、分支限界法中的至少一种并结合大数据平台构建服务模型。
8. 一种智能服务大数据建模系统,其特征在于,所述系统包括:
业务原始数据库获取单元,用于获取业务原始数据库,所述原始数据库包括各个安防要素数据;
安防主题库获取单元,用于根据使用场景要求对原始数据库中的安防要素进行特征提取,形成若干安防主题库;
数据服务目录构建单元,用于构建数据服务目录;

安防服务开放单元,用于将数据服务目录汇聚到大数据平台,进行安防服务开放。

9.一种计算机存储介质,其特征在于:所述计算机存储介质存储有多条指令,所述指令适于由处理器加载并执行如权利要求1-7任一项所述的方法步骤。

10.一种电子设备,其特征在于,包括处理器和存储器,所述存储器存储有计算机程序,所述计算机程序适于由所述处理器加载并执行如权利要求1-7任一项所述的方法步骤。

智能服务大数据建模方法、系统、存储介质及计算机设备

技术领域

[0001] 本实施例涉及安防系统大数据服务技术领域,尤其涉及智能服务大数据建模方法、系统、存储介质及计算机设备。

背景技术

[0002] 目前安防信息管理应用的各类数据资源分散,碎片化严重,目标人员管控等缺乏工具支撑。在技术方面,安防云及建模平台、大数据滴灌行动、省市政府部门间信息共享、大数据平台、网侦数据的深化应用、“四标多实”管控要素的采集应用等提供了应用和数据支撑。已具备建设统一协作的智慧平台的基础条件。其中在打防管控方面,目前一般是通过以往工作经验来判断,无法感知监测各类风险,做到情报信息灵敏,精准评估预警。没有相关的研判平台和数据建模,没有系统的监测手段,还无法高效的解决社区管理中的突出问题,解决群众反映严重的问题,也无法为人们提供智能化、便捷化、人性化的服务,切实提高群众的获得感、满意度。

发明内容

[0003] 本发明提供了一种智能服务大数据建模方法、系统、存储介质及计算机设备,提升打防管控和内部管理水平,依托大数据建模,提升研判分析能力,拓展智慧民生,提升民生服务水平。

[0004] 本发明的技术方案为:

[0005] 第一方面,本申请提供了一种智能服务大数据建模方法,其特征在于,包括:

[0006] 获取业务原始数据库,所述原始数据库包括各个安防要素数据;

[0007] 根据使用场景要求对原始数据库中的安防要素进行特征提取,形成若干主题库;

[0008] 构建数据服务目录;

[0009] 将数据服务目录汇聚到大数据平台,进行安防服务开放。

[0010] 进一步地,获取业务原始数据库的方法包括:

[0011] 获取各个离散的、多元多态的独立文件并汇聚形成原始数据项;

[0012] 对所述原始数据项通过大数据平台进行清洗加工,获得所述业务原始数据库。

[0013] 进一步地,所述大数据平台通过PL/SQL Developer、TOAD或Navicat Premium进行数据清洗加工,对数据进行清洗加工的方法包括:

[0014] 对数据进行规则定义,具体包括对数据进行基础信息的定义、数据映射关系的定义、数据分级分类的定义和数据质量核验规则的定义中的至少一种操作;

[0015] 和/或根据关键字段进行数据提取;

[0016] 和/或根据数据定义对数据进行数据过滤、去重、格转、效验中的至少一种操作;

[0017] 和/或还包括将数据通过分治法、动态规划法、贪心算法、回溯法、分支限界法中的至少一种算法与其他相关业务对应的进行关联回填、关联提取、关联分析;

[0018] 和/或与大数据平台的数据进行结构化和/或非结构化对比。

[0019] 进一步地,所述主题库包括目标人员主题库、场所主题库、物品主题库、案件主题库、事件主题库、组织主题库。

[0020] 进一步地,构建数据服务目录的方法包括:

[0021] 根据不同的业务应用需要并按照分层解耦的方法构建数据服务目录;

[0022] 根据数据服务目录中设置满足业务直接调用集成的不同的数据服务接口,可通过不同的数据服务接口调用不同的业务应用。

[0023] 进一步地,所述数据服务接口还设置有访问规则和权限,方便进行数据鉴权,数据鉴权的维度包括内容敏感度、数据来源、数据种类、字段及字段关系分类。

[0024] 进一步地,安防服务包括将安防数据按照数据敏感级别分成不同等级,进行分级授权,还包括提供模型分析及数据推送服务,其中,模型分析服务包括根据业务需要,利用分治法、动态规划法、贪心算法、回溯法、分支限界法中的至少一种并结合大数据平台构建服务模型。

[0025] 第二方面,本申请还提供了一种智能服务大数据建模系统,其特征在于,

[0026] 所述系统包括:

[0027] 业务原始数据库获取单元,用于获取业务原始数据库,所述原始数据库包括各个安防要素数据;

[0028] 安防主题库获取单元,用于根据使用场景要求对原始数据库中的安防要素进行特征提取,形成若干主题库;

[0029] 数据服务目录构建单元,用于构建数据服务目录;

[0030] 安防服务开放单元,用于将数据服务目录汇聚到大数据平台,进行安防服务开放。

[0031] 第三方面,本申请还提供了一种计算机存储介质,所述计算机存储介质存储有多条指令,所述指令适于有处理器加载并执行上述方法步骤。

[0032] 第四方面,本申请还提供了一种计算机设备,包括处理器和存储器,所述存储器存储有计算机程序,所述计算机程序适于由所述处理器加载并执行上述方法步骤。

[0033] 本申请一些实施例提供的技术方案带来的有益效果至少包括:通过获取包括各个安防要素数据,根据使用场景要求对原始数据库中的要素进行特征提取,形成若干主题库,然后构建数据服务目录,最后将数据服务目录汇聚到大数据平台,进行安防服务开放,实现了内外部资源的汇聚,提升了打防管控和内部管理水平,同时依托大数据建模,提升了研判分析能力,拓展了智慧民生,提升了民生服务水平。

附图说明

[0034] 为了更清楚地说明本发明实施例,下面将对实施例或现有技术描述中所需要使用的附图做简单介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,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。

[0035] 图1为本发明实施例的方法流程图;

[0036] 图2为本发明实施例的方法流程框图;

[0037] 图3为本发明实施例的系统框图;

[0038] 图4为本发明实施例的装置的硬件结构图。

[0039] 以下结合附图对本发明的具体实施方式作进一步详细地说明。

具体实施方式

[0040] 下面结合附图对本发明进行详细描述,在详述本发明实施例时,为便于说明,表示器件结构的附图会不依一般比例做局部放大,而且所述示意图只是示例,其在此不应限制本发明保护的范围。需要说明的是,附图采用简化的形式且均使用非精准的比例,仅用以方便、清晰地辅助说明本发明实施例的目的。同时,在本申请的描述中,术语“第一”、“第二”等仅用于区分描述,而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量;术语“正”、“反”、“底”、“上”、“下”等指示的方位或位置关系为基于附图所示的方位或位置关系,仅是为了便于描述本发明和简化描述,而不是指示或暗示所指的装置或元件必须具有特定的方位、以特定的方位构造和操作,因此不能理解为对本发明的限制。

[0041] 在本发明的描述中,需要说明的是,除非另有明确的规定和限定,术语“相连”、“连接”应做广义理解,例如,可以是固定连接,也可以是可拆卸连接,或一体地连接;可以是机械连接,也可以是电连接;可以是直接相连,也可以通过中间媒介间接相连,还可以是两个元件内部的连通,可以是无线连接,也可以是有线连接。对于本领域的普通技术人员而言,可以具体情况理解上述术语在本发明中的具体含义。

[0042] 需要说明的是,本申请实施例提供的一种智能服务大数据建模方法一般由计算机设备执行,相应的,智能服务大数据建模系统一般设置于计算机设备中。

[0043] 本申请的计算机设备上可以安装有各种通讯客户端应用,例如:视频录制应用、视频播放应用、语音交互应用、搜索类应用、即时通讯工具、邮箱客户端、社交平台软件等。

[0044] 计算机设备可以是硬件,也可以是软件。当计算机设备为硬件时,可以是具有显示屏的各种计算机设备,包括但不限于智能手机、平板电脑、膝上型便携式计算机和台式计算机等等。当计算机设备为软件时,可以是安装上述所列举的计算机设备中。其可以实现呈多个软件或软件模块(例如:用来提供分布式服务),也可以实现成单个软件或软件模块,在此不作具体限定。

[0045] 当计算机设备为硬件时,其上还可以安装有显示设备和摄像头,显示设备显示可以是各种能实现显示功能的设备,摄像头用于采集视频流;例如:显示设备可以是阴极射线管显示器(cathode ray tube display,简称CR)、发光二极管显示器(light-emitting diode display,简称LED)、电子墨水屏、液晶显示屏(liquid crystal display,简称LCD)、等离子显示面板(plasma display panel,简称PDP)等。用户可以利用计算机设备上的显示设备,来查看显示的文字、图片、视频等信息。

[0046] 本实施例主要依托大数据平台提供的数据,创建安防部分通用的、基础的数据模型和服务,通过模型输出安防数据的结果,同时可以实现反哺至大数据平台,补充安防数据模型。

[0047] 如图1所示,下面以具体实施例来介绍本申请的一种智能服务大数据建模方法,所述方法包括:

[0048] 首先,获取业务原始数据库,所述原始数据库包括各个安防要素数据。

[0049] 为满足安防平台的业务应用目标实现,需要汇聚多个其他业务系统中所需的安防要素数据,如图2所示,安防要素数据有安防云数据、警情数据、安防服务数据、小区感知数

据等等,可以使用专业的数据接入工具,例如其他共用大数据平台的数据接入工具,保证源业务系统数据的历史全量和新增数据稳定,做到及时汇聚,在本实施例中,数据可以统一汇聚到具备数据收集及处理功能的智慧安防平台。

[0050] 汇聚的数据称为原始数据项,由于源业务系统建设的独立性、数据标准不统一、业务场景不统一,数据的格式、数据项存在格式不准确、重复、空白、数据含义不一等问题,为便于数据使用,需要使用数据处理工具对数据进行清洗加工,形成业务原始库。

[0051] 以上列举的三个软件程序都可以集成在大数据平台上进行初步的数据清洗处理。

[0052] 对数据进行清洗的方法可以包括以下几个方面:

[0053] 第一,可以结合业务需求在起始阶段定义数据接入、处理的规则。

[0054] 数据资源基础信息处理规则:提供数据资源基础信息的建立、修改、删除等操作。可定义数据组织分类、数据资源名称、来源地区、应用系统等基础信息。

[0055] 数据资源接入定义:提供配置对普通文档进行定义数据处理的过程,对原始文本分隔符、换行符进行批量转换和处理、对不同编码的原始文本进行格式转化的建立、修改、删除等功能。

[0056] 数据映射定义:对数据映射的建立、修改、删除等操作定义。数据映射定义的内容包括两个方面,一是原始字段项与标准数据元映射关系的定义,二是原始字典代码集与规范化字典代码集映射关系的定义。

[0057] 数据分级分类管理:定义数据项的字段性质的分类,对数据项新建、修改、删除等操作,及定义数据项的字段敏感度分类。

[0058] 数据质量核检规则管理设定:依据《数据质量管理技术规范》中的评估要求,管理数据资源的质量核检规则的建立、修改、删除等,并对数据质量报表分析统计。

[0059] 设定好规则后,可采用与大数据平台兼容的数据处理工具,如PL/SQL Developer, TOAD,Navicat Premium等,进行数据清洗加工。

[0060] 其中,PL/SQL Developer:是一个集成开发环境,由Allround Automations公司开发,专门面向Oracle数据库存储的程序单元的开发。有越来越多的商业逻辑和应用逻辑转向了Oracle Server,因此,PL/SQL编程也成了整个开发过程的一个重要组成部分。PL/SQL Developer侧重于易用性、代码品质和生产力,充分发挥Oracle应用程序开发过程中的主要优势的;

[0061] TOAD:是一套专业的数据库开发工具,透过图形式的操作接口,大幅提升程序开发人员或数据库管理员的作业效率。超强功能、超低成本的特点,让程序开发人员运用TOAD开发PL/SQL时,更为简便、迅速,同时也有效减低数据库管理员的工作负荷,TOAD支持的数据库包括:Oracle、DB2、SQLServer、MySQL;

[0062] Navicat Premium:是一个可多重连接的数据库管理工具,它可让你以单一程序同时连接到MySQL、MariaDB、MongoDB、SQL Server、Oracle数据库、PostgreSQL和SQLite数据库,让管理不同类型的数据库更加方便。Navicat Premium结合了其他Navicat成员的功能。有了不同数据库类型的连接能力,Navicat Premium支持在MySQL、MariaDB、MongoDB、SQL Server、Oracle数据库、PostgreSQL和SQLite数据库之间传输数据。它支持大部份MySQL、MariaDB、MongoDB、SQL Server、Oracle数据库、PostgreSQL和SQLite的功能;

[0063] 第二,根据数据定义结果进行数据过滤、去重、格转、校验等操作,生成满足标准及

质量要求的数据。数据清洗是实现数据标准化的主要环节。

[0064] 例如格转：身份证15位转18位：

[0065] 编码规则：公民身份号码是特征组合码，由十七位数字本体码和一位校验码组成。排列顺序从左至右依次为：六位数字地址码，八位数字出生日期码，三位数字顺序码和一位校验码，可以用字母表示如为ABCDEFYYMMDDXXXR。其含义如下：

[0066] 1. 地址码(ABCDEF)：表示编码对象常住户口所在县(市、旗、区)的行政区划代码，按GB/T2260的规定执行。

[0067] 2. 出生日期码(YYMMDD)：表示编码对象出生的年、月、日，按GB/T7408的规定执行，年、月、日分别用4位、2位(不足两位加0)、2(同上)位数字表示，之间不用分隔符。

[0068] 3. 顺序码(XXX)：表示在同一地址码所标识的区域范围内，对同年、同月、同日出生的人编定的顺序号，顺序码的奇数分配给男性，偶数分配给女性。

[0069] 4. 校验码(R)，一位数字，通过前17位数字根据一定计算得出。

[0070] 校验码是通过一系列数学计算得出来的，具体校验的计算方式如下：

[0071] 1. 对前17位数字本体码加权求和；

[0072] 公式为： $S = \text{Sum}(A_i * W_i), i = 0, \dots, 16;$

[0073] 其中 A_i 表示第 i 位置上的身份证号码数字值， W_i 表示第 i 位置上的加权因子，其各位对应的值依次为：7 9 10 5 8 4 2 1 6 3 7 9 10 5 8 4 2通俗解释：身份证第一位数字 X_7 +第二位 X_9 +第三位 X_{10} +第四位 X_5 +第五位 X_5 +第六位 X_8 +第七位 X_4 +第八位 X_1 +第九位 X_6 +第十位 X_3 +十一位 X_7 +十二位 X_9 +十三位 X_{10} +十四位 X_5 +十五位 X_8 +十六位 X_4 +十七位 X_2 ；计算出总和(用S)表示。

[0074] 2. 以11对计算结果取模；

[0075] $Y = \text{mod}(S, 11);$

[0076] 通俗解释：用S除以11，看最后的余数。如果除尽，为0；余数为1，则计为1；最大余数为10；全部数字为0-10共11个数字。(用Y表示)。

[0077] 3. 根据模的值得到对应的校验码；

[0078] 对应关系为：

[0079] Y值：0 1 2 3 4 5 6 7 8 9 10

[0080] 校验码：1 0X 9 8 7 6 5 4 3 2

[0081] 通俗解释：余数为0，则校验码为1；依次类推：余数为1，则校验码对应0；以下：2--X；3--9；4--8；5--7；6--6；7--5；8--4；9-3；10-2。

[0082] 如果校验码不符合这个规则，则肯定是假号码。

[0083] 过滤：通过对信息进行辨别和分离，实现冗余及垃圾信息的滤除。基于数据标准和过滤规则，对不合标准及规则的数据过滤；基于样本和内容分析，对冗余或垃圾信息进行辨别、分离和过滤。被识别为冗余或垃圾信息的数据可以直接滤除，或标识后照常处理并交由后端模块(或人工)判断如何进一步处理。

[0084] 去重：在各类场景下设定相应的数据重复判别规则以及合并、清除策略，对数据进行重复性辨别，并对重复数据进行合并或清除处理。

[0085] 格转：根据数据元标准把非标数据转换成统一的标准格式进行输出。针对不同来源的同类数据按照统一规则进行转换，如数据字典、地址门牌、公民身份号码、手机号、IP地

址、时间、经纬度等属性的标准格式转换。

[0086] 校验:根据校验知识库对数据进行检验,符合标准的数据直接入库,不符合标准的数据可进入问题数据库以便进一步分析处理。校验主要包括数据的完整性校验、一致性校验等。常用的校验规则有:空值校验、取值范围校验、公民身份号码/手机号/车牌号/IMEI/MAC/IP地址等校验、数值校验、长度校验、精度校验等。此外,还有更为复杂的多字段条件校验、业务规则校验等。

[0087] 第三,根据数据定义中的关联规则或算法(分治法、动态规划法、贪心算法、回溯法、分支限界法),可以调用大数据平台的数据关联支撑服务,将数据和其它知识数据、业务数据等进行关联,并输出关联信息,还可以进行业务库表数据的关联回填、关联提取、关联分析。

[0088] 1>、关联回填

[0089] 关联回填是将不完备的日志数据与知识数据等按场景进行关联,并将关联的要素等信息回填到日志,提升数据的关联及价值。

[0090] 人员身份关联回填:根据数据内容,回填人员的详细身份信息,如:根据银行账户信息,回填开户人的身份证号信息。

[0091] 车辆物品关联回填:针对包含车辆信息的日志数据,回填车辆详细信息,如:根据车牌号和车牌号类型,回填车主信息。

[0092] 时空关联回填:针对轨迹、行为数据,回填空间信息,如:在旅店住店信息中回填旅店的地址、经纬度信息。

[0093] 手机账号关联回填:针对互联网日志数据,根据接入认证信息回填手机认证账号信息。

[0094] imsi账号关联回填:针对互联网日志数据,根据接入认证信息回填imsi认证账号信息。

[0095] 宽带账号关联回填:针对互联网日志数据,根据接入认证信息回填宽带认证账号信息。

[0096] 手机注册信息关联:针对包含手机号的日志数据,回填手机号的注册信息,如:根据手机号信息,回填手机号的注册机主姓名、身份证号信息。

[0097] 活动场所关联:对活动场所信息进行回填,根据活动场所编码,回填活动场所的名称、地址、经纬度信息。

[0098] 关联举例,如车辆物品关联回填:

[0099] 通过车辆轨迹信息表中的车牌号码与车主信息表中的车牌号码进行关联,获取车主信息表中的信息,如车主姓名,联系电话,可以回填至车辆轨迹信息中,车辆轨迹信息表中的车牌号类型可以回填至车主信息表中。

[0100] 2>、关联提取

[0101] 关联提取,是根据提取规则,对各类数据资源中涉及的关键要素关系或关联进行提取,主要包括:要素关联信息提取、要素关系信息提取、民族信息关联提取。

[0102] 第四,可以按照规则对结构化和非结构化数据进行进行相同比较或相似度计算,并跟大数据平台的数据做比对,对于命中规则的数据,按照输出描述进行输出,用于信息布控和信息订阅等。例如:

[0103] 1>、结构化比对

[0104] 通过对线索(如身份证件号码、护照号、车辆、银行卡、手机号、终端特征、网络身份等)的比对,在海量日志数据中命中发现线索相关信息。

[0105] 需要说明的是,完全匹配是检索线索内容与比对字段内容完全相同。

[0106] 模糊匹配是线索内容在比对字段内容中出现,则匹配成功。

[0107] 范围匹配是线索内容与指定的字段进行对比,指定的字段内容在线索内容区间,则匹配成功。比如某个坐标系范围内的比对。

[0108] 正则匹配是线索内容为正则表达式,指定的字段内容符合线索内容设定的规则,则匹配成功。

[0109] 2>、非结构化比对

[0110] 关键词比对:通过对关键词及关键词组合的比对,在海量全文数据中命中发现关键词相关信息。

[0111] 文本相似度比对:对线索文本进行特征抽取,与原始数据中的文本数据进行文本特征比对,返回相似度数值结果、对应文本信息。

[0112] 二进制比对:对二进制线索文件进行MD5值计算,与非结构化数据的MD5值进行比对,返回相似度数值结果、对应非结构化数据。

[0113] 多媒体特征比对:语种比对、图像比对、视频比对、生物特征(如指纹、面部特征、声纹、虹膜、DNA等)的比对。

[0114] 需要说明的是,对数据进行清洗处理加工的实施服务,主要针对第一次接入的源业务数据系统开展工作,建立固定的数据处理规则、模板、元数据定义之后,以后数据处理就可自动标准化处理,该源业务数据系统的新增数据就可以自动加工,确保数据结果标准化,无需长期的数据清洗工作。除非通过数据管理工具,如共用大数据平台的数据管理工具,发现数据质量存在变化时,会终止数据读取再次进行源业务数据系统的数据探查分析。

[0115] 第五,根据业务需要,利用分治法、动态规划法、贪心算法、回溯法、分支限界法等,结合大数据平台,开发数据统计、分析、规律性探索及预测等,建立按照服务种类分类的异常分析类、关系分析类、群体分析类、内容分析类、人物画像类、轨迹分析行为分析类等分析服务模型,以支撑应用层业务场景复杂、多变的需求。模型分析服务还可按照功能划分为预测类服务、数据集碰撞、分析服务等。

[0116] 数据集碰撞类服务:根据条件中需要碰撞的字段,在一个或多个数据集中比对,按照交集、并集或差集进行计算,并返回结果。

[0117] 分析类服务:使用各种分析方法分析数据,以获取数据的统计分布情况,发现数据的内在规律性、识别其主要因素,或进行模型的参数估计、可信度评估等。

[0118] 预测类服务:利用各类已有模型、算法对数据进行计算,预测未知的变量或属性取值。

[0119] 需要说明的是,分治法是把一个复杂的问题分成两个或更多的相同或相似的子问题,再把子问题分成更小的子问题……直到最后子问题可以简单的直接求解,原问题的解即子问题的解的合并。

[0120] 动态规划法是每次决策依赖于当前状态,又随即引起状态的转移。一个决策序列就是在变化的状态中产生出来的,所以,这种多阶段最优化决策解决问题的过程就称为动

态规划。

[0121] 贪心算法是在对问题求解时,总是做出在当前看来是最好的选择。也就是说,不从整体最优上加以考虑,他所做出的仅是在某种意义上的局部最优解。常见的贪心算法有:Prim算法、Kruskal算法(都是求最小生成树的)。

[0122] 回溯法实际上一个类似枚举的搜索尝试过程,主要是在搜索尝试过程中寻找问题的解,当发现已不满足求解条件时,就“回溯”返回,尝试别的路径。

[0123] 深度优先;

[0124] 分支限界法类似于回溯法,也是一种在问题的解空间树T上搜索问题解的算法。但在一般情况下,分支限界法与回溯法的求解目标不同。回溯法的求解目标是找出T中满足约束条件的所有解,而分支限界法的求解目标则是找出满足约束条件的一个解,或是在满足约束条件中找出使某一目标函数值达到极大或极小的解,即在某种意义下的最优解。

[0125] 第六,数据推送设置,设置包括数据推送配置、数据封装及数据推送的数据推送配置。

[0126] 其中,数据推送配置:包括推送规则、推送标准、数据种类、推送周期等配置。

[0127] 数据封装:根据配置进行数据封装。

[0128] 数据推送:将数据资源推送给目标平台或系统。

[0129] 另外,列举说明本实施例的自动加工方法,当数据的规则定义都设定好后,新数据就可以按照这些规则进行自动加工了,如证件号码:

[0130] 数据源字段:

[0131] 常住人口信息表中证件号码,在进行自动加工时,需要编写程序包体,进行证件号码的确认,确认规则如下:如1代表公民身份证号码,规则为身份证号码为18位,前17位不能有字母存在,必须为纯数字,18位中前两位代表省份,每个身份编码对应数据字典表,如果发现包括字母的,或者不在数据字典表中,那么进行自动跳过;第三,四位代表城市,同理,第五六位代表区域,与上述同理;第七位到第十位代表年份,有效年份为正则表达式进行校验,目前值在1900--2099年,超过或者小于均为无均数据,会自动跳过;第十一、十二位为月份,有效值为01-12,超过或小于均为无效值,同样跳过,第十三十四位为日期,其中要判断月份为02时,是否是闰年,如果是闰年,值在01-29,如果是非闰年,值在01-28,最后第十五位至第十七位为数字代表顺序,最后一位为数字0-9或字母X,有且只有是X,大小写均可。如果以上任何一个条件不满足,那么数据均会自动跳过。数据自动加工为每天凌晨1-2点开始。

[0132] 然后,根据使用场景要求对原始数据库中的安防要素进行特征提取,形成若干主题库。

[0133] 需要说明的是,应按照使用场景要求进行数据组织,智慧安防平台核心是安防对象管理,安防对象管理的核心是对象与社会接触要素的管理,为此在数据处理过程中,对安防对象要素特征进行提取、不同要素关系和关联关系确定、按照知识规则对数据进行打标签,支持安防对象管控的实时数据布控需求等,形成身份证号、MAC地址、车牌号、人脸等不同要素的业务专题库,也就是业务要素库。

[0134] 在本实施例中,主题库可按照人、车、MAC地址、场所、组织、案件、事件等对象感知进行主题库划分。主题库可划分为目标人员主题库、场所主题库、物品主题库、案件主题库、

事件主题库、组织主题库,进行对象数据的有价值、提前归拢。这样业务应用就是在比较高效的数据组织下进行开发和应用,而不是直接基于数据源不够稳定的原始数据进行应用开发,符合现代业务应用开发技术路线要求。

[0135] 业务主题库是为了便利工作、精准快速地反映工作对象全貌而建立的融合各类内务管理、业务管理、协作管理、研判分析等的主题库,长期积累形成的多种维度的公共数据集合,具体的,在本实施例中:

[0136] 目标人员主题库:人员是目标人员所涉及到的自然人信息的高度抽象和归类,从大数据平台抽取相关信息,补全业务系统和感知系统的轨迹行为数据;

[0137] 人员基本信息主要包括:自然信息情况(姓名、性别等)、地址(籍贯、出生地等)、生物特征情况(身高、近视等)、参加或创建的群组或组织等等。

[0138] 人员证照信息主要包括:驾驶证、行驶证、居住证、暂住人口、签证的基本信息和照片信息等。

[0139] 人员背景信息主要包括:教育经历、社会经历、工作经历、培训经历等。

[0140] 人员资产信息主要包括:车辆、房产、工商、税务、水电、银行卡等信息。

[0141] 人员活动轨迹主要包括:航班、铁路、手机、硬件特征串等活动轨迹、活动规律、常出现地点等信息。

[0142] 人员行为信息主要包括:网吧、宾旅馆、寄递、消费、通话、网上活动等信息。

[0143] 人员涉案信息主要包括:涉及事件网上舆情事件,涉及网下案件情况违法情况。

[0144] 人员关系信息主要包括:血缘关系、婚姻关系等。

[0145] 场所主题库:是指对安防辖区内特定建筑物或公共空间活动处所的多维刻画,主要包括:银行、出租房屋、公共复杂场所、寺庙、歌舞厅、重点单位、电站、电视台、小区等。

[0146] 物品主题库:物品主题库是安防所关注重点对象的所涉及物品信息的高度抽象和概括,并按物品所固有的特征加以表述,包含但不限于以下范围:

[0147] 身份标识物:银行卡、电话号码、虚拟身份(上网账号,及社交、购物、游戏等平台账号)、驾驶证、社会保障卡、DNA、指纹等;

[0148] 交通工具:火车、飞机、普通汽车、专用汽车、摩托车等;

[0149] 计算机设备及产品:便携机、网卡、硬盘等;

[0150] 通信设备:手机、传真机、交换机、对讲机等;

[0151] 案件主题库:事件主题库是对安防辖区内事件的发端、演化、处置以及涉及事件的相关人、地、物、组织等多维刻画。

[0152] 案件主题库从案件安防信息、案件基本信息、案件立案信息、案件单位信息及调查处置信息、案件目标人员信息及调查、措施等信息、案件相关的网上线索、案件涉物品信息及调查处置信息、结案信息、移送处置信息、案件情报信息及流转处置信息、案件其他相关人信息等几个维度来进行设计。

[0153] 事件主题库:事件主题库是安防辖区内所发生事件的发端、演化、处置以及涉及事件的相关人、地、物、组织等多维刻画。

[0154] 组织主题库:组织主题库是对安防辖区内的如非重点关注组织的多维刻画。包含但不限于以下范围:

[0155] 单位:事业(包括行政单位等)、企业;

- [0156] 社团:老乡会、商会、学生社团、学术社团、行业社团、联合社团、专业社团;
- [0157] 一般群体:户籍(家庭、集体)、合租;
- [0158] 以上主题库是通过大数据平台进行数据抽取汇聚至本地数据库中,根据各主题库要求,进行所有表的字段的清洗和拆分,形成各主题库内容。
- [0159] 还需要说明的是,可以通过select语句对各相关字段进行提取,例如在提取姓名时主要是提取姓名等关键、基础、重要信息。
- [0160] 例如:常住人口表中提取姓名字段信息;
- [0161] `Insert into xm value as Select name from za_czrk,`
- [0162] `'za_czrk',sysdate;`
- [0163] 向姓名表中插入常住人口信息表中的姓名字段信息,同时插入数据来源为“za_czrk”及创建时间;
- [0164] 提取地址:主要是提取地址等关键、基础、重要信息。
- [0165] 例如:常住人口表中提取常住地址字段信息;
- [0166] `Insert into dz value as Select czdz from za_czrk, 'za_czrk',sysdate;`
- [0167] 向地址表中插入常住人口信息表中的常住地址字段信息,同时插入数据来源为“za_czrk”及创建时间;
- [0168] 提取身份证号:主要是提取身份证等关键、基础、重要信息。
- [0169] `Insert into zjhm value as Select zjhm,'身份证' from za_czrk, 'za_czrk',sysdate;`向地址表中插入常住人口信息表中的证件号码字段和“身份证”标识信息,同时插入数据来源为“za_czrk”及创建时间;
- [0170] 提取手机号:主要是提取手机号码等关键、基础、重要信息。
- [0171] `Insert into sjh value as Select lxdh from za_czrk, 'za_czrk',sysdate;`
- [0172] 向地址表中插入常住人口信息表中的联系电话字段信息,同时插入数据来源为“za_czrk”及创建时间。
- [0173] 采用同样的方法,根据需求设定提取的规则对数据进行提取。
- [0174] 提取之后,构建数据服务目录;
- [0175] 在本步骤中,应遵循分层解耦建设本平台所需的数据服务目录。
- [0176] 具体的,通过数据服务接口实现应用、数据的分层解耦。例如,采用包含安防业务库的数据查询、数据操作(DML操作)、数据分析计算、数据比对等数据服务接口,比如常口库、流口库等提供的数据接口,及集成大数据平台提供的背景审查、风险评估等数据服务接口,并且按照数据服务目录方式进行统一的组织管理,满足业务应用直接调用集成的需要。通过数据服务接口也便于控制数据授权和鉴权,鉴权时通过采用访问规则和权限,进行数据鉴权,数据鉴权服务是基于数据的访问控制规则,实现数据的访问权限鉴别的过程。访问控制规则从内容敏感度、数据来源、数据种类、字段及字段关系分类四个维度进行资源权限的控制,资源鉴权通过用户的数据资源权限,使用数据鉴权服务实现对数据资源的访问控制。
- [0177] 最后,将数据服务目录汇聚到大数据平台,进行安防服务开放。
- [0178] 智慧安防平台数据可以实时汇聚到到大数据平台,按照分级分类进行安防服务开放。目前,结合大数据平台的安防系统、视频系统、110系统、人口系统等部门系统间进行数

据交换和信息推送,进而实现数据汇聚和数据下发,还可以将数据资源根据需从下级大数据中心往上级数据中心汇聚,或者从外部单向导入,汇到相应的各级数据中心的数据汇聚、将数据资源根据需从部省级数据中心下发给下级数据中心数据下发等。在信息推送时,可以根据服务模型进行对象发现、对象刻画、对象评估、对象干预方面的预警推送。

[0179] 本实施例方法可以将数据及时全量汇聚到大数据平台,作为大数据平台的区域数据的主要来源之一。进入大数据平台之后,与其他更多的数据,尤其是网络数据进行再次融合治理,虚实关联,丰富画像特征,支持开发更多智能性服务和应用,再反哺智慧安防平台。安防数据会按照数据敏感级别分成不同等级,根据使用环境、使用者的岗位职责与案事件级别等综合性因素,进行分级授权,还可结合人员的业务经验、知识、授权规则进行决策辅助服务。

[0180] 第二方面,如图3所示,本申请实施例还提供了一种智能服务大数据建模系统,所述智能服务大数据建模系统可为上述智慧安防平台或集成在上述智慧安防平台上,所述系统包括:

[0181] 业务原始数据库获取单元201,用于获取业务原始数据库,所述原始数据库包括各个安防要素数据;

[0182] 安防主题库获取单元202,用于根据使用场景要求对原始数据库中的安防要素进行特征提取,形成若干主题库;

[0183] 数据服务目录构建单元203,用于构建数据服务目录;

[0184] 安防服务开放单元204,用于将数据服务目录汇聚到大数据平台,进行安防服务开放。

[0185] 需要说明的是,上述实施例提供的系统在执行上述方法时,仅以上述各功能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模块完成,即将设备的内部结构划分成不同的功能模块,以完成上述的全部或者部分功能。另外,上述实施例提供的系统与方法实施例属于同一构思,其体现实现过程详见方法实施例,这里不再赘述。

[0186] 上述本申请实施例序号仅仅为了描述,不代表实施例的优劣。

[0187] 本申请实施例还提供了一种计算机存储介质,所述计算机存储介质存储有多条指令,所述指令适于有处理器加载并执行上述智能服务大数据建模的方法步骤。

[0188] 具体执行过程可以参见图1所示实施例的具体说明,在此不进行赘述。

[0189] 本申请实施例还提供了一种计算机设备,包括处理器和存储器,所述存储器存储有计算机程序,所述计算机程序适于由所述处理器加载并执行上述智能服务大数据建模的方法。

[0190] 请参见图4,为本申请实施例提供了智能服务大数据建模装置的主要硬件结构示意图。如图4所示,

[0191] 智能服务大数据建模装置300可以包括:至少一个处理器301,至少一个网络接口304,用户接口303,存储器305,至少一个通讯总线302。

[0192] 其中,通讯总线302用于实现这些组件之间的连接通讯。

[0193] 其中,用户接口303可以包括触摸屏和摄像头(Camera)。

[0194] 其中,网络接口304可选的可以包括标准的有线接口、无线接口(如WI-FI接口)。

[0195] 其中,处理器301可以包括一个或者多个处理核心。处理器301利用各种接口和线路连接整个智能服务大数据建模装置300内的各个部分,通过运行或执行存储在存储器305内的指令、程序、代码集或指令集,以及调用存储在存储器305内的数据,执行智能服务大数据建模装置300的各种功能和处理数据。可选的,处理器301可以采用数字信号处理(Digital Signal Processing,DSP)、现场可编程门阵列(Field-Programmable Gate Array,FPGA)、可编程逻辑阵列(Programmable Logic Array,PLA)中的至少一种硬件形式来实现。处理器301可集成中央处理器(Central Processing Unit,CPU)、图像处理器(Graphics Processing Unit,GPU)和调制解调器等中的一种或几种的组合。其中,CPU主要处理操作系统、用户界面和应用程序等;GPU用于负责显示屏所需要显示的内容的渲染和绘制;调制解调器用于处理无线通讯。可以理解的是,上述调制解调器也可以不集成到处理器301中,单独通过一块芯片进行实现。

[0196] 其中,存储器305可以包括随机存储器(Random Access Memory,RAM),也可以包括只读存储器(Read-Only Memory)。可选的,该存储器305包括非瞬时性计算机可读介质(non-transitory computer-readable storage medium)。存储器305可用于存储指令、程序、代码、代码集或指令集。存储器305可包括存储程序区和存储数据区,其中,存储程序区可存储用于实现操作系统的指令、用于至少一个功能的指令(比如触控功能、声音播放功能、图像播放功能等)、用于实现上述各个方法实施例的指令等;存储数据区可存储上面各个方法实施例中涉及到的数据等。存储器305可选的还可以是至少一个位于远离前述处理器301的存储装置。如图4所示,作为一种计算机存储介质的存储器305中可以包括操作系统、网络通讯模块、用户接口模块以及应用程序。在图4所示的系统300中,用户接口303主要用于为用户提供输入的接口,获取用户输入的数据;而处理器301可以用于调用存储器305中存储的应用程序,并具体执行如图1所示的方法,具体过程可参照图1所示,此处不再赘述。

[0197] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,所述的程序可存储于一计算机可读取存储介质中,该程序在执行时,可包括如上述各方法的实施例的流程。其中,所述的存储介质可为磁碟、光盘、只读存储记忆体或随机存储记忆体等。

[0198] 对所公开的实施例的上述说明,使本领域专业技术人员能够实现或使用本发明。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的,本文中所定义的一般原理可以在不脱离本发明的精神或范围的情况下,在其它实施例中实现。因此,本发明将不会被限制于本文所示的实施例,而是要符合与本文所公开的原理和新颖特点相一致的最宽范围。

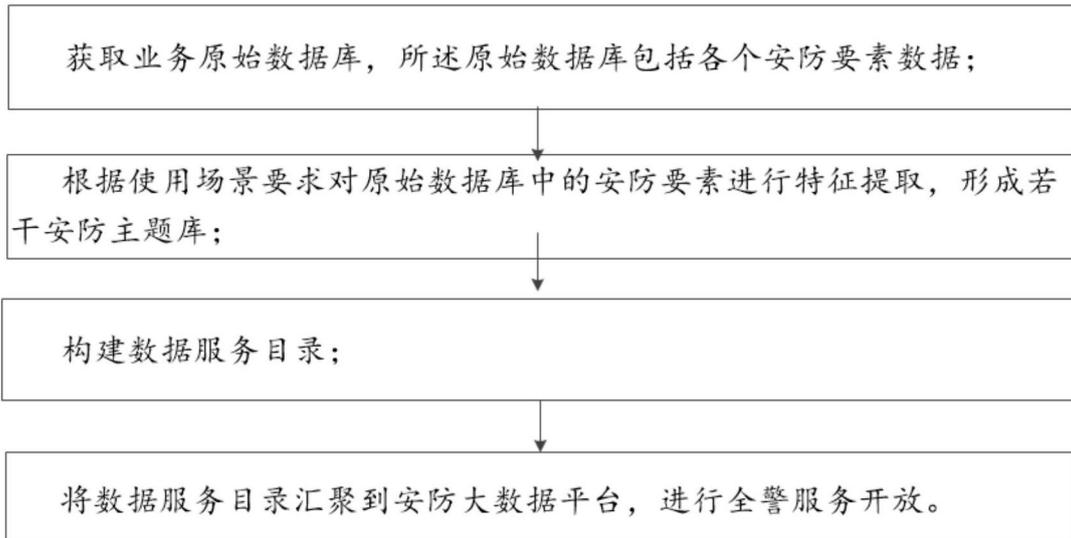


图1

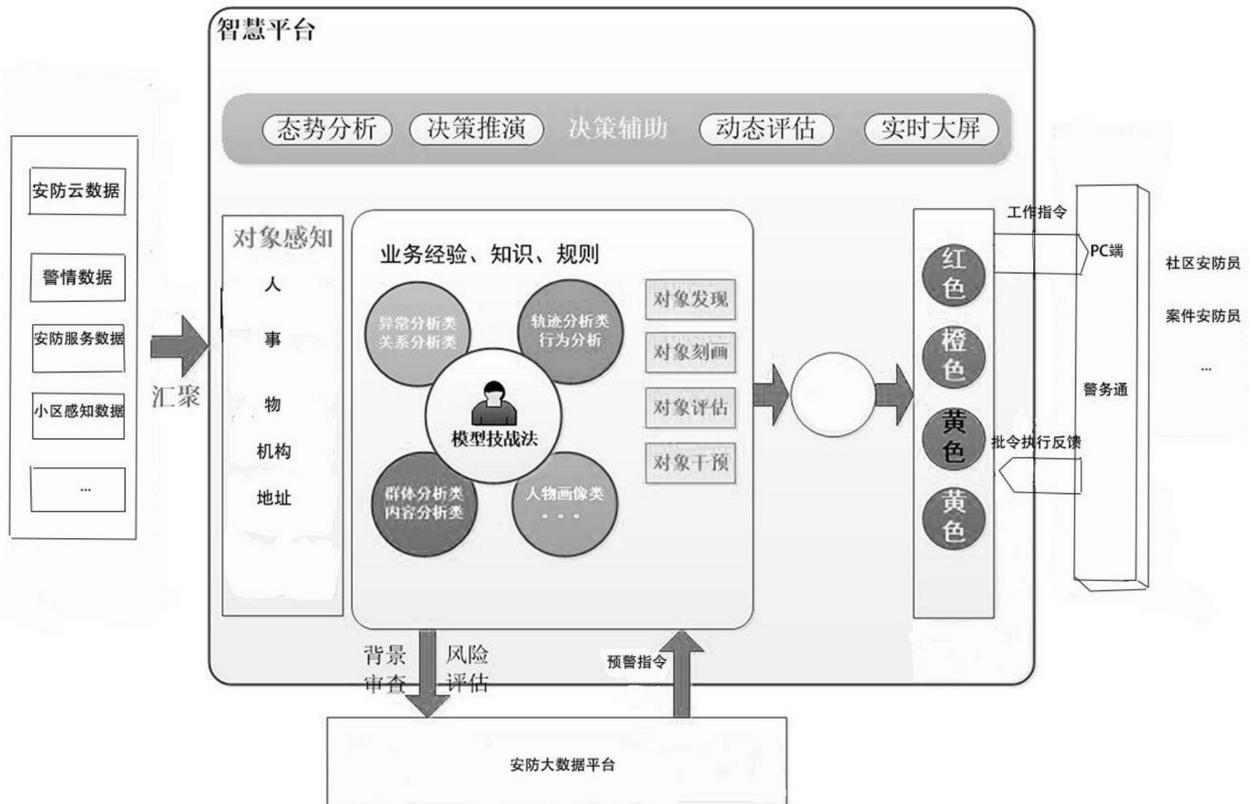


图2

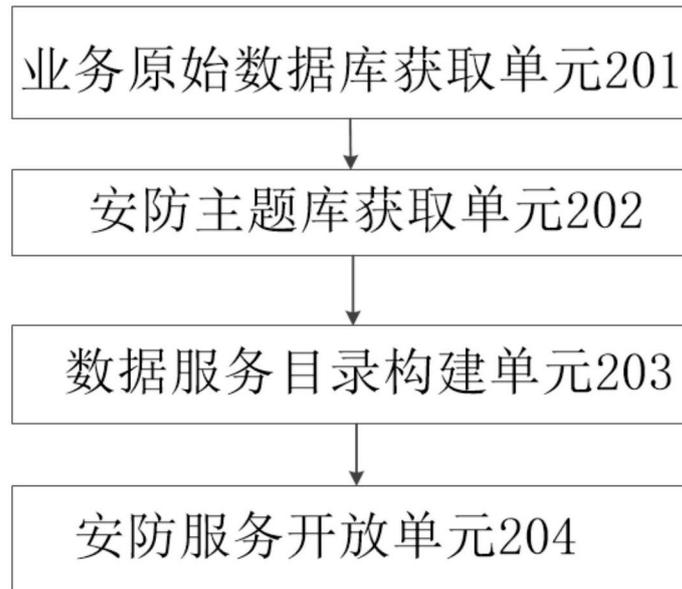


图3

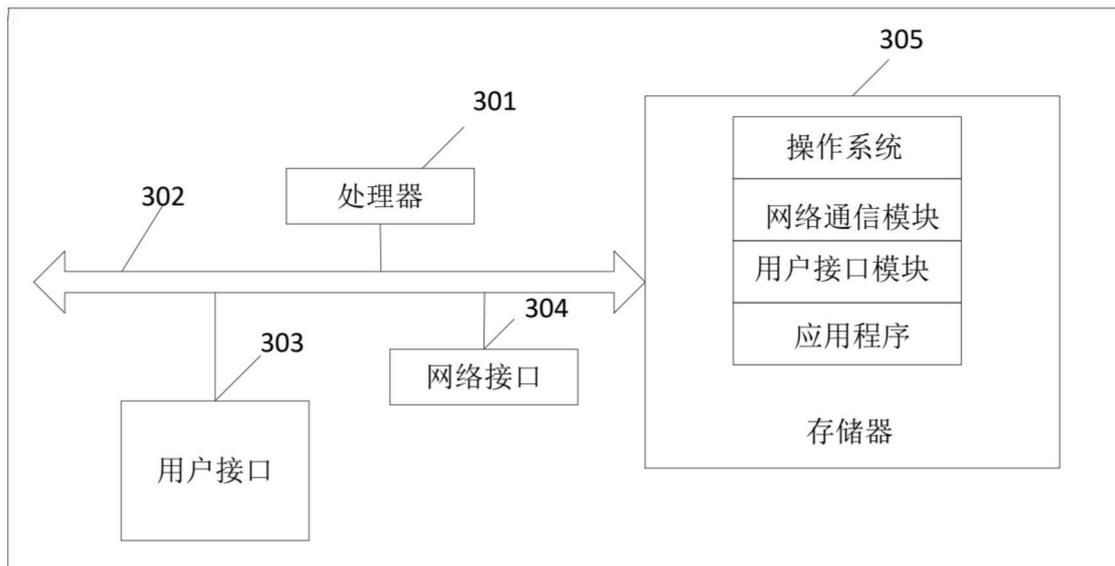


图4