

基于辐射关系的数据行为响应机制研究

李春生, 李少龙, 邸京华, 张可佳

(东北石油大学 计算机与信息技术学院, 黑龙江 大庆 163318)

摘要:在数据集成应用于构建智能化平台过程中,随着规模的不断扩大,大量数据的高频变更行为带来的联动影响导致平台工作稳定性及效率出现瓶颈。为了提高平台处理数据联动变更业务的准确性和高效性,文中提出数据辐射模型概念,借鉴分类分析方法描述辐射行为,针对其复杂程度与辐射范围制定出完整的行为响应规则。同时为进一步优化行为反馈效率,缩短处理时间,设计出基于时间代价优化的RD-TCO模型用于处理数据辐射行为,并完成基于辐射关系的数据行为响应机制(RDBR),以达到增加数据处理准确度、提高数据集成平台工作效率的目的。最后结合大庆油田某采油厂三次采油跟踪运行数据管理平台的实际使用情况,利用RDBR机制实现对平台中数据集成系统的设计与改进。

关键词:数据辐射;行为响应;时间代价优化;RD-TCO模型

中图分类号:TP312

文献标识码:A

文章编号:1673-629X(2016)06-0146-05

doi:10.3969/j.issn.1673-629X.2016.06.032

Research on Response Mechanism of Data Behavior Based on Radiation Relationship

LI Chun-sheng, LI Shao-long, DI Jing-hua, ZHANG Ke-jia

(College of Computer and Information Technology, Northeast Petroleum University, Daqing 163318, China)

Abstract: In the process of applying data integration in building an intelligent platform, with the continuous expansion of scale, the high frequency change of large amounts of data has caused a bottleneck in the work stability and efficiency of the platform. In order to improve accuracy and effectiveness in data linkage model, a radiation relationship is proposed and referring to classification analysis method, rules of behavior response for different situations is formulated, drawing up a complete response rule based on its complexity and radiation range. At the same time, the RD-TCO model is designed to optimize the feedback efficiency and shorten time cost of response, finishing the response mechanism of data behavior based on radiation relationship. To achieve the purpose that increase the accuracy of data and improve the efficiency of the data integration platform. Finally, in combination with the three oil recovery operation platform of a production plant in Daqing Oilfield, the design of data integration system in the platform via RDBR mechanism is implemented.

Key words: data radiation; behavioral response; optimization of time cost; RD-TCO model

0 引言

数据集成技术经过数十年的发展,已经广泛应用于工业生产及工程施工各领域中,并发挥着举足轻重的作用^[1]。由于数据规模的扩大和开发时效性的增强,对相关生产数据准确性与高效性的需求与日俱增,于是探究数据内部联动关系,制定迅捷有效的数据行为响应机制成为提高集成平台工作效率的关键^[2-4]。辐射关系的提出将数据联动影响规则化,降低了大量数据行为变化过程中来源缺失、关系变更等现象带来的负面影响,解决了数据集成平台工作过程中由于数

据联动关系带来的冗余与模糊^[5]。

集成平台对数据动态的分析与管理为企业提供了实时的数据视图与灵活的查询服务^[6]。虽然通过制定基于标准的元数据模型实现了数据变更的协同性,但由于集成数据频繁的行为变更与复杂的关系网络而依旧存在以下不足:

(1) 由于工程需求的多变性,导致部分数据缺失状况严重或使用价值较小,造成数据来源定义模糊,数据获取方式混乱^[2,7]。

(2) 部分有动态化需求的数据由于关联数据从属

收稿日期:2015-09-11

修回日期:2015-12-11

网络出版时间:2016-05-25

基金项目:黑龙江省科学基金项目(F2015020)

作者简介:李春生(1960-),男,博士,教授,博士生导师,研究方向为人工智能及其应用、模式识别与人工智能;李少龙(1990-),男,硕士研究生,通讯作者,研究方向为人工智能与信息处理技术。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20160525.1706.032.html>

关系的变更,计量单位、语义属性、计算方式等因素的改变,极大程度地降低了数据的准确度。

(3)当多个数据变更行为发生时,彼此之间涉及的数据模型可能存在错综复杂的获取顺序问题,导致集成平台工作效率的低下^[8]。

针对上述问题,设计基于辐射关系的数据行为响应机制,提出数据辐射模型表示发生行为变化数据的关系属性及逻辑结构。通过分类分析的方法构建数据行为响应规则,设计 RD-TCO 模型,实现对多行为并发响应顺序的优化,达到提高数据联动变更准确率与多变更行为处理效率的目的。

1 引入数据辐射模型

数据辐射概念源于数据集成平台工作中频发的数据变更过程所产生的联动影响,现将其定义为由于部分数据模型发生逻辑属性、计算方式、数值内容等数据行为变化造成关联数据模型属性与功能的变更^[9-10]。

在数据集成平台中,当部分数据发生行为变化时,称该数据为辐射源数据 D_s 。 D_s 是造成数据辐射的主要原因。将受 D_s 影响发生改变的数据集合称为辐射域 β , β 内具体受影响的数据元素称为辐射域影响因子 φ 。

数据辐射模型可描述为一个五元组^[11](触发标识 γ , 辐射元素 Rd_f , 业务属性 Bu_f , 辐射域 β , 辐射关系 α)。

触发标识 γ 作为检测数据是否发生变更行为的触发器,默认未发生变更时状态为 0,发生变更时置为 1。辐射元素 Rd_f 是模型中发生变化的数据元素集合,具体表示形式为:

$$Rd_f = \{r_1, r_2, \dots, r_k\}$$

业务属性 Bu_f 用以描述 Rd_f 中数据元素在实际项目中代表的含义以及计量单位等具体功能。辐射域 β 表示受 Rd_f 辐射的数据模型的集合矩阵,具体表示形式如下:

$$\beta = \begin{pmatrix} r_1 & Bu_{f1} & \alpha_1 \\ \vdots & \vdots & \vdots \\ r_n & Bu_{fn} & \alpha_n \end{pmatrix}$$

辐射关系 α 包含了辐射数据 Rd_f 与辐射域 β 中对应数据元素的获取方式。因此一个基于辐射关系的源数据模型可直接表示为:

$$D = \{\gamma, Rd_f, Bu_f, \beta, \alpha\}$$

2 辐射关系的分类

辐射关系分为数值行为辐射与关系行为辐射^[12]。其中数值行为辐射较为简单,只需考虑数值变化时对

辐射数据的影响,具体包括:

数值变更。源数据模型中只有数据元素的数值发生变化,其一般表述形式为:

$$D_0 = \{\gamma, Rd_{f0}, Bu_f, \beta, \alpha\} \rightarrow$$

$$D_n = \{\gamma, Rd_{fn}, Bu_f, \beta, \alpha\}$$

业务属性变更。在项目中表达的含义发生变化,如月度数据更改为旬度数据、计量单位的变化等,可能会造成数值的连带变更,其一般表述形式为:

$$D_t = \{\gamma, Rd_{ft}, Bu_{ft}, \beta, \alpha_k\} \rightarrow$$

$$D_n = \{\gamma, Rd_{fn}, Bu_{fn}, \beta, \alpha\}$$

关系行为辐射因源数据与辐射数据对应关系发生变化需要考虑因素众多,较为复杂,具体包括如下两种情况。

(1)辐射关系变更。源数据模型获得辐射数据时对应的计算方式发生变化,其一般表述形式为:

$$D_k = \{\gamma, Rd_{fk}, Bu_{fk}, \beta, \alpha_k\} \rightarrow$$

$$D_n = \{\gamma, Rd_{fn}, Bu_{fn}, \beta, \alpha_n\}$$

(2)辐射源变更。辐射源的变更分为新增、删除与替换每项变更在发生时都有可能造成对应关系的变化,其一般表述形式为:

$$D_p = \{\gamma, Rd_{fp}, Bu_{fp}, \beta_p, \alpha_p\} \rightarrow$$

$$D_n = \{\gamma, Rd_{fn}, Bu_{fn}, \beta_n, \alpha_n\}$$

在分析辐射源变更情况时,首先对源数据与辐射数据的对应关系进行分类。

(1)SCI(Single Correlation Line)型关系。对于一个源数据模型 M_1 和一个辐射数据模型 M_2 ,引入表示两者关联变化的函数 $\text{link}(\delta)$,其一般表述形式为:

$$r = \text{link}(e) \{e \in Rd_1 \in M_1 \mid r \in Rd_2 \in M_2\}$$

当且仅当 $e \rightarrow e_n$ 时总有:

$$r_n = \text{link}(e_n)$$

则认为 M_1 和 M_2 存在 SCI 型关系。

(2)SCS(Single Correlation Star)型关系。对于一个源数据模型 M_u 和多个辐射数据模型 $\{M_1, M_2, \dots, M_k\}$,引入表示多重对应关系关联变化的函数 $\text{link}_u(\delta)$,存在

$$\begin{cases} r_1 = \text{link}_{u1}(e), r_1 \in \beta_1 \in M_1 \\ r_2 = \text{link}_{u2}(e), r_2 \in \beta_1 \in M \\ \vdots \\ r_k = \text{link}_{uk}(e), r_k \in \beta_1 \in M_k \end{cases}, (e \in Rd_1 \in M_u)$$

当 $e \rightarrow e_n$ 时总有:

$$\begin{cases} r_{n1} = \text{link}_{u1}(e_n) \\ r_{n2} = \text{link}_{u2}(e_n) \\ \vdots \\ r_{nk} = \text{link}_{uk}(e_n) \end{cases}$$

则认为 M_1 和 M_2, M_3, \dots, M_k 存在 SCS 型关系。

(3) MCN (Multiple Correlation Net) 型关系。对于多个源数据模型 M_1, M_2, \dots, M_t 和一个辐射数据模型 M_v , 引入表示多个辐射来源对应关系的关联变化函数 $\text{link}_p(\delta)$, 存在:

$$r_v = \text{link}_p(e_1, e_2, \dots, e_t); \{e_1 \in \text{Rd}_1 \in M_1, e_2 \in \text{Rd}_2 \in M_2, \dots, e_t \in \text{Rd}_t \in M_t\}$$

当 $e_1 \rightarrow e_{n1}, e_2 \rightarrow e_{n2}, \dots, e_t \rightarrow e_{nt}$ 时总有:

$$r_{vn} = \text{link}_p(e_{n1}, e_{n2}, \dots, e_{nt})$$

则认为 M_1, M_2, \dots, M_t 和 M_v 存在 MCN 型关系。由于辐射数据模型中不同数据元素可以受多个辐射源影响, 所以辐射源与辐射数据的对应关系可以是三种关系的交集。

行为响应规则的制定如下所述:

针对数值行为和关系行为具体分类情况, 研究辐射数据在受到数据行为影响后的变化规律, 提出如下响应规则:

Rule1: 当发生辐射行为时, 将触发标识 γ 置为 1, 并判断辐射行为如果是数值行为则进行 Rule2, 否则进行 Rule3。

Rule2: 当发生数值变更行为时, 依据原有辐射关系搜索辐射域影响数据元素和业务属性进行修改。

Rule3: 判断关系变更行为类型, 如果是辐射关系的变更则进行 Rule4, 否则进行 Rule5。

Rule4: 当发生辐射关系变更时, 修改辐射关系, 将新的辐射关系 α_n 赋予源数据模型。

Rule5: 当变更行为是辐射源的增加时, 建立新的辐射关系。如果变更行为是辐射源的删除且关系不为 MCN 型, 则将辐射域 β 内所有辐射数据元素 Rd_f 删除, 否则判断其他 $n-1$ 个辐射源是否发生辐射关系之内的替换行为, 发生则对源数据模型的辐射域和辐射关系进行修改, 不发生则直接进行删除。

当变更行为是辐射源的替换时, 将原有辐射关系 α 从旧辐射源中剔除, 如果是 SCI 型、SCS 型, 则关系 α 赋予新辐射源, 否则建立新辐射关系。

Rule6: 在 Rule2、Rule4、Rule5 完成之后将辐射数据模型标记为新的辐射源模型, 重新进行 Rule1 直至触发标识为 0。

3 基于 RD-TCO 模型的响应执行过程

通过对辐射关系的分类以及针对性的行为响应规则设计, 对数据辐射行为有了基本的处理方法。但当多个辐射行为发生时, 数据模型通常会处于一个较为复杂的关系网之中, 因此设计一套高效有序的响应执行模型变得至关重要^[13]。将辐射关系简化为一个执行顺序有向图 G , D 表示辐射关系中所有涉及的数据

模型集合, F 表示执行辐射响应所需时间代价的集合, 因此行为响应模型的实际功能转化为图的路径搜索优化问题。涉及路径问题常使用的几种算法其时间代价相对固定, 且不能直观地反映各顶点之间路径序列的先后关系^[14]。所以, 为了提高效率, 减少无谓的重复比较计算, 文中提出了基于数据行为响应顺序的时间代价优化算法。

设计数据行为响应顺序的基本思想可以认为是建立起当前状态和之前状态的一种转移表示, 通过分析两种状态之间的差别来确定执行顺序的优化程度。用一个 n 阶矩阵 T 来表示图的任意两点最短路径, $T[i, j]$ 表示点 i 到点 j 最短路径的长度, $L(i, j)$ 表示边的权值即辐射关系所需时间代价。 T 的初始值如下:

$$T[i, j] = \begin{cases} 0, & i = j \\ L(i, j), & i \neq j, (i, j) \in E \\ \infty, & i \neq j, (i, j) \notin E \end{cases}$$

T 不一定是要求的最短路径, 在某些简单情况下, T 可能与最后要求解的最短路径矩阵恰好相等。 $T[i, j]$ 值所表达的含义是点 i 直接到点 j 的权值, 即 $L(i, j)$, 当 i 不能到达 j 时其权值为 ∞ , 用 $T^{(k)}[i, j]$ 表示在 $\{1, 2, \dots, k\}$ 限制下的一般路径算法的状态转移方程 $T^k[i, j]$:

$$\min(T^{(k-1)}[i, j], T^{(k-1)}[i, k] + T^{(k-1)}[k, j])$$

通过对方程的观察, 发现 n 次状态转移过程中, 每次转移都要枚举所有点对点 (i, j) 的路径, 所以时间代价固定为 $O(n^3)$ 。但在实际计算过程中, 中间节点的插入很多时候不能使路径变短, 因此基于优化时间代价的角度, 算法具体步骤为:

第一步, 作初始状态矩阵 $T^{(0)} = (t_{ij}^{(0)})$ 和关系矩阵 $R^{(0)} = (r_{ij}^{(0)})$, 其中

$$t_{ij}^{(0)} = \begin{cases} L(i, j), & (i, j) \in E \\ \infty, & (i, j) \notin E \end{cases}$$

$$r_{ij}^{(0)} = \begin{cases} 1, & (i, j) \in E \\ f, & (i, j) \notin E \end{cases}$$

第二步, 构造时间代价矩阵 $T^{(k)} = (t_{ij}^{(k)})$ 和关系矩阵 $R^{(k)} = (r_{ij}^{(k)})$ 。对于 $T^{(k)}$ 中元素 $t_{ij}^{(k)}$, 当节点 V_r 插入时, 如果 $t_{ir}^{(k-1)} \cup t_{rj}^{(k-1)} \geq t_{ij}^{(k-1)}$, 说明插入节点 V_r 后路径长度不变, 此时无需计算 $t_{ir}^{(k-1)} + t_{rj}^{(k-1)}$ 。否则:

$$t_{ij}^{(k)} = \min\{t_{ij}^{(k-1)}, t_{ir}^{(k-1)} + t_{rj}^{(k-1)}\}$$

同时, 若 $t_{ij}^{(k)} = t_{ir}^{(k-1)} + t_{rj}^{(k-1)}$, 则记下点 V_1 并将关系矩阵中 $R^{(k)}$ 对应的元素 $r_{ij}^{(k)}$ 变为:

$$r_{ij}^{(k)} = \{r_{ir}^{(k-1)}, V_r, r_{rj}^{(k-1)}\}$$

第三步, 若 $T^{(k+1)} = T^{(k)}$, 得到最优执行顺序, 算法

终止,否则返回第二步。

RD-TCO 模型的引入如下所示。

应用时间代价优化算法进行 RD-TCO 模型的构建。对图 1 所示的执行顺序有向图 G_1 进行执行顺序求解。

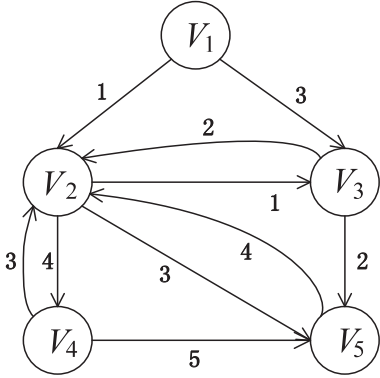


图 1 执行顺序有向图 G_1

针对数值行为和关系行为具体分类情况,构建初始状态矩阵 $\mathbf{T}^{(0)}$ 和关系矩阵 $\mathbf{R}^{(0)}$:

$$\mathbf{T}^{(0)} = \begin{bmatrix} 0 & 1 & 3 & \infty & \infty \\ \infty & 0 & 1 & 4 & 3 \\ \infty & 2 & 0 & \infty & 2 \\ \infty & 3 & \infty & 0 & 5 \\ \infty & 4 & \infty & \infty & 0 \end{bmatrix}$$

$$\mathbf{R}^{(0)} = \begin{bmatrix} 1 & 1 & 1 & f & f \\ f & 1 & 1 & 1 & 1 \\ f & 1 & 1 & f & 1 \\ f & 1 & f & 1 & 1 \\ f & 1 & f & f & 1 \end{bmatrix}$$

研究辐射数据在受到数据行为影响后的变化规律,提出响应规则。参照初始状态矩阵 $\mathbf{T}^{(0)}$ 和关系矩阵 $\mathbf{R}^{(0)}$ 计算 $\mathbf{T}^{(1)}$ 和 $\mathbf{R}^{(1)}$ 。由于 $t_{11}^{(0)} = 0$,故 $t_{11}^{(1)}$ 不需计算,直接得到 $t_{11}^{(1)} = t_{11}^{(0)} = 0$,与此对应的关系矩阵 $\mathbf{R}^{(1)}$ 的 $r_{11}^{(1)} = r_{11}^{(0)} = 0$ 保持不变。

对 $t_{12}^{(1)}$ 进行分析, $\mathbf{T}^{(0)}$ 中第一行除 $t_{11}^{(1)}$ 之外所有的值都不小于 $t_{12}^{(0)}$,易得 $t_{12}^{(1)} = t_{12}^{(0)} = 1$,相应的关系矩阵 $\mathbf{R}^{(0)}$ 的 $r_{12}^{(1)} = r_{12}^{(0)} = 1$ 保持不变。

再对 $t_{13}^{(1)}$ 进行分析。由于 $\mathbf{T}^{(0)}$ 中第一行除 $t_{11}^{(1)}$ 之外比 $t_{13}^{(0)}$ 小的值只有 $t_{12}^{(1)}$,固有:

$$t_{13}^{(1)} = \min\{t_{13}^{(0)}, t_{12}^{(1)} + t_{23}^{(0)}\}$$

对关系矩阵 $\mathbf{R}^{(0)}$:

$$r_{13}^{(1)} = \{r_{12}^{(1)}, V_2, r_{23}^{(0)}\} = \{V_2\}$$

此时将节点 V_2 加入路径,即 V_1 到 V_3 的路径需经过 V_2 。

依据上述分析方法, $t_{14}^{(1)}$ 和 $t_{15}^{(1)}$ 计算结果如下:

$$t_{14}^{(1)} = \min\{t_{14}^{(0)}, t_{12}^{(1)} + t_{24}^{(0)}\}$$

$$t_{15}^{(1)} = \min\{t_{15}^{(0)}, t_{12}^{(1)} + t_{25}^{(0)}, t_{13}^{(1)} + t_{35}^{(0)}\}$$

插入节点 V_2 加入路径:

$$r_{14}^{(1)} = \{r_{12}^{(1)}, V_2, r_{24}^{(0)}\} = \{V_2\}$$

$$r_{15}^{(1)} = \{r_{12}^{(1)}, V_2, r_{25}^{(0)}\} = \{V_2\}$$

将方法推广至 $\mathbf{T}^{(1)}$ 和 $\mathbf{R}^{(1)}$ 中所有元素,得到时间代价矩阵 $\mathbf{T}^{(1)}$ 和关系矩阵 $\mathbf{R}^{(1)}$:

$$\mathbf{T}^{(1)} = \begin{bmatrix} 0 & 1 & 2 & 5 & 4 \\ \infty & 0 & 1 & 4 & 3 \\ \infty & 2 & 0 & 6 & 2 \\ \infty & 3 & 4 & 0 & 5 \\ \infty & 4 & 5 & 8 & 0 \end{bmatrix}$$

$$\mathbf{R}^{(1)} = \begin{bmatrix} 1 & 1 & V_2 & V_2 & V_3 \\ f & 1 & 1 & 1 & 1 \\ f & 1 & 1 & V_2 & 1 \\ f & 1 & V_2 & 1 & 1 \\ f & 1 & V_2 & V_2 & 1 \end{bmatrix}$$

显然有 $\mathbf{T}^{(1)} \neq \mathbf{T}^{(0)}$,需要继续执行顺序优化,利用上述方法得到时间代价矩阵 $\mathbf{T}^{(2)}$ 和关系矩阵 $\mathbf{R}^{(2)}$:

$$\mathbf{T}^{(2)} = \begin{bmatrix} 0 & 1 & 2 & 5 & 4 \\ \infty & 0 & 1 & 4 & 3 \\ \infty & 2 & 0 & 6 & 2 \\ \infty & 3 & 4 & 0 & 5 \\ \infty & 4 & 5 & 8 & 0 \end{bmatrix}$$

$$\mathbf{R}^{(2)} = \begin{bmatrix} 1 & 1 & V_2 & V_2 & V_3 \\ f & 1 & 1 & 1 & 1 \\ f & 1 & 1 & V_2 & 1 \\ f & 1 & V_2 & 1 & 1 \\ f & 1 & V_2 & V_2 & 1 \end{bmatrix}$$

经观察比较得知 $\mathbf{T}^{(2)} = \mathbf{T}^{(1)}$,算法结束。所以 $\mathbf{T}^{(2)}$ 中对应的元素 $t_{ij}^{(2)}$ 的值即为节点 V_i 和节点 V_j 之间的最短路径。

通过这一处理,算法的时间复杂度降低了,尤其是当图是稀疏图时,一般传统算法依然要 $O(n^3)$,改进后算法的用时几乎为零。设指向 i 点的入点有 u_i 个,且由 i 指出的出点有 v_i 个,则显然有:

$$\sum_{i=1}^n u_i = \sum_{i=1}^n v_i = p, \quad 0 \leq p \leq n(n-1)$$

其中, p 表示图的稀疏程度。

时间 T 可以表示为:

$$T = n \left[\sum_{i=1}^n u_i v_i \right] \approx \frac{\sum_{i=1}^n u_i \sum_{i=1}^n v_i}{n} = \frac{p^2}{n}$$

由此可知,时间的复杂度为 $O(p^2/n)$,当图为稀疏图,即 $p \rightarrow 0$ 时,则时间复杂度 $O(p^2/n) \rightarrow 0$ 。当需响应的数据行为数量较少时,模型在效率上的优势不

易察觉,当变更样本容量逐步扩大,复杂度逐步增加时,RD-TCO 模型比一般传统方法时间代价上的缩短变得尤为明显,如表 1 所示。

表 1 RD-TCO 模型与一般传统方法性能比较表

分组	N	U	V	语句执行频度	
				一般传统方法	RD-TCO
简单	100	1	50	$10^4 + 10^6$	1.5×10^4
	100	50	50	$10^4 + 10^6$	2.6×10^5
	100	99	99	$10^4 + 10^6$	9.9×10^5
复杂	1 000	1	500	$10^6 + 10^9$	1.5×10^6
	1 000	500	500	$10^6 + 10^9$	2.51×10^8
	1 000	999	999	$10^6 + 10^9$	9.91×10^8

由此可见,基于时间代价优化的 RD-TCO 模型在复杂行为关系执行顺序的问题上具有高效紧凑的优势^[15]。

4 设计实例

以 RD-TCO 模型为基础,设计数据辐射行为响应机制,对油田施工中的生产数据及相关作业历史数据的联动变更进行更新与修改,应用在大庆油田某采油厂地质大队、作业大队等相关单位。选择较为有代表性的采油区块月度注入数据进行说明,如表 2 所示。

表 2 采油区块月度注入动态数据表

辐射因子名称	辐射源	辐射关系
开井率 KJL	总井数、开井数	开井数/总井数
注入压力 ZRYL	注入井月数据	当月井史开井注入压力平均值
允注压差 RZYC	区块基础数据、注入压力	破裂压力与注入压力的差值
...
视吸水指数 SXSZS	区块基础数据、注入量、注入压力、开井数	注入量/注入压力/开井数/区块基础数据的有效厚度
注采比 ZCB	采出动态数据、注入量	注入量/(采出动态数据的井口日产液+井口日产油*0.31)

表 2 选取了区块月度注入动态数据模型中若干数据元素,既包括受多个辐射源影响的视吸水指数、注采比,还有模型内部辐射的开井率等,因此极具代表性。为了提高效果的直观性,选取 6 个变更样本,如表 3 所示。

根据辐射类型将样本进行分类,可知 1、2 号样本属于数值变更,5 号属于业务属性变更,3 号属于辐射关系变更。4、5 号分别属于 MCN 型、SCI 型辐射源变更。

分析 6 个样本的辐射关系和涉及数据元素估算具

体行为执行时间代价,根据 RD-TCO 模型计算出最优执行顺序为:

$$V_1 \rightarrow V_2$$
$$V_1 \rightarrow V_4 \rightarrow V_3$$
$$V_1 \rightarrow V_4$$
$$V_1 \rightarrow V_4 \rightarrow V_3 \rightarrow V_5$$

经过上述分析计算可知,当辐射行为的并行发生量增至 1 000 条时,传统方法较基于辐射关系的行为响应机制语句执行频度高出约 56%,所以该机制对于数据执行效率具有较为明显的提升。

表 3 数据行为变更样本表

序号	样本变更内容	涉及数据元素
1	去除注入井月数据中井号 N2-5-229、N2-50-P2231、N2-4-P220 的数据	开井数、开井率、视吸水指数
2	井号 N8-4-26 的油价降低	注入压力、允注压差、视吸水指数
3	允注压差计算方式变为:取区块基础数据的破裂压力与注入压力差值的 80%	允注压差
4	视吸水指数计算方式变为(注入量/注入压力/开井数/区块基础数据的砂岩厚度)	视吸水指数
5	分注井划分方式变为当月井史注水方式等于 122	分注井

将上述样本实验应用于另外 20 余个辐射关系模型,其中 17 个模型在实际运行过程中未出现系统崩溃和数据错误问题,2 个模型运行过程中出现系统崩溃,经工作人员排查,系计算机硬件问题,1 个模型发生数据错误问题,属于人工录入错误。由此可见基于辐射关系的行为响应机制可以应用于数据集成平台等领域。

根据基于辐射关系的行为响应机制开发的数据管理平台已经正常工作,凭借对联动数据变更高效而准确的处理受到了应用单位的较好评价。

5 结束语

文中提出了基于辐射关系的数据行为响应机制,定义了数据辐射基本概念,对辐射关系中数据元素进行结构化的描述,采用分类分析的方法对数据行为进行了分类并设计了针对分类情况的响应规则,同时根据多辐射关系特征引入基于时间代价优化算法的 RD-TCO 模型完成了行为响应机制并节约了响应的时间成本,实现了提高集成数据联动反应准确度与效率的目的。

种瓶盖区域提取的方法。在瓶盖区域中寻找瓶盖圆心,以定位飞边位置,采用图像差分法计算飞边面积,利用加权色差公式查出色差,该系统在实际检测过程中表现出了较高的准确性与稳定性。算法在实际应用中,需根据所检测的瓶盖尺寸以及颜色做出对应的参数调整,并配合生产线上的逻辑电路控制拍照与剔除缺陷瓶盖等自动化操作。

参考文献:

- [1] 陈靖宇,付江云,谢振南,等. 基于实时滤波的瓶盖密封性测试系统设计[J]. 计算机测量与控制,2012,20(3):951-954.
- [2] 郑云峰,王贵锦,何 贝,等. 基于直线拟合算法的 PET 瓶瓶盖检测方法[J]. 计算机应用研究,2011,28(11):4398-4400.
- [3] Zhou Wenju, Fei Minrui, Zhou Huiyu, et al. A sparse representation based fast detection method for surface deflection of bottle caps[J]. Neurocomputing, 2014, 123(10):406-414.
- [4] 郑海成,谢 云. PET 瓶封盖质量在线检测系统的设计与实现[J]. 电子技术应用,2012,38(8):20-22.
- [5] 刘 杰,安博文. 基于动态阈值分割的目标提取技术[J]. 红外技术,2008,30(12):706-708.
- [6] 陈 磊. 视频图像中的运动目标检测算法研究[D]. 长沙:湖南大学,2007.
- [7] Lv Zhenhan, Lin Liqun, Li Yuan. A PET bottle mouth detection method based on morphologic[C]//Proc of third international joint conference on computational science and optimization.

(上接第150页)

参考文献:

- [1] Mezni H, Chainbi W, Ghedira K. AWS-Policy: an extension for autonomic web service description[J]. Procedia Computer Science, 2002, 10:915-920.
- [2] 刘 威,杨 丹. 基于虚拟视图的异构数据库集成平台的研究[J]. 计算机技术与发展,2009,19(6):91-94.
- [3] 宋洪涛. 基于 SDO 的异构数据集成系统的设计与实现[D]. 北京:北京邮电大学,2009.
- [4] 丰江帆,李林,杨富平,武志涛,袁正午. 基于 Service GIS 的震害应急数据集成方法研究与应用[J]. 重庆邮电大学学报:自然科学版,2011,23(1):111-114.
- [5] 徐俊刚,裴 莹. 数据 ETL 研究综述[J]. 计算机科学,2011,38(4):15-20.
- [6] 丁卫亮. 基于本体异构数据集成方法的研究[D]. 杭州:浙江工业大学,2009.
- [7] 徐 晶,许 炜. 消息中间件综述[J]. 计算机工程,2005,31(16):73-76.
- [8] 曹洋洋. 空间数据库增量联动更新办法研究[D]. 杭州:浙

江大学,2010.

- [8] Shen Yehu, Mo Rui, Wei Lei. Bottle cap scratches detection with computer vision techniques[C]//Proc of ninth international conference on natural computation. [s. l.]: [s. n.], 2013.
- [9] Ramli S, Mustafa M M, Hussain A, et al. Automatic detection of 'ROIs' for plastic bottle classification[C]//Proc of the 5th student conference on research and development. Malaysia: [s. n.], 2007.
- [10] 田 巍,庄镇泉. 基于 HSV 色彩空间的自适应肤色检测[J]. 计算机工程与应用,2004,40(14):81-85.
- [11] 马 玲,张晓辉. HSV 颜色空间的饱和度与明度关系模型[J]. 计算机辅助设计与图形学学报,2014,26(8):1272-1278.
- [12] Gonzalez R C, Woods R E. Digital image processing[M]. 阮秋琦,阮宇智,译. 第3版. 北京:电子工业出版社,1998.
- [13] 朱芳芳,顾宏斌,孙 瑾. 一种改进的 Hough 变换直线检测算法[J]. 计算机技术与发展,2009,19(5):19-22.
- [14] 吴慧兰,刘国栋,刘炳国,等. 基于 SIFT 算法的圆心快速精确定位技术研究[J]. 光电子·激光,2008,19(11):1512-1515.
- [15] 陈 毅,刘晓玉,蒋 峥,等. 基于 RGB 和 HSV 的胶囊异囊缺陷识别方法[J]. 计算机工程与设计,2014,35(11):3888-3892.
- [16] Pavlova P E, Cyrilov K P, Moumdjiev I N. Application of HSV colour system in identification by colour of biological objects on the basis of microscopic images[J]. Computerized Medical Imaging and Graphics, 1996(5):357-364.

江大学,2014.

- [9] 林 源,陈志泊. 分布式异构数据库同步系统的应用[J]. 计算机工程与设计,2010,31(24):5278-5281.
- [10] 黄建平. 基于 SQL Server 数据库日志的信息源监测方法的研究与实现[D]. 广州:暨南大学,2007.
- [11] 高雅田,李春生,富 宇. 基于关系数据分析的知识服务模式[J]. 计算机工程,2011,37(5):56-58.
- [12] 张 锋,张莉莉. 触发器在数据处理过程中的应用研究[J]. 计算机工程与科学,2008,30(5):156-158.
- [13] Hu J, Khalil I, Han S, et al. Seamless integration of dependability and security concepts in SOA: a feedback control system based framework and taxonomy[J]. Journal of Network and Computer Applications, 2011, 34:1150-1159.
- [14] 徐小玲,彭 京,石葆梅,等. 一种基于边序列的任意两点间最短路径算法[J]. 计算机工程与应用,2005,41(29):88-90.
- [15] 袁鼎荣,张师超,朱晓峰,等. 基于相对等待时间的代价敏感决策树[J]. 计算机科学与探索,2007(3):314-324.