



(12) 发明专利申请

(10) 申请公布号 CN 122195969 A

(43) 申请公布日 2026. 06. 12

(21) 申请号 202610210727.5

G06F 16/29 (2019.01)

(22) 申请日 2026.02.13

G06F 16/22 (2019.01)

(71) 申请人 广东省林业调查规划院

地址 510520 广东省广州市天河区广汕一路338号

申请人 广东省城乡规划设计研究院科技股份有限公司

(72) 发明人 秦琳 吴晓生 蔡伟 张水花

刘新科 林寿明 陈志敬 陈得冠 鲁好君

(74) 专利代理机构 广州粤高专利商标代理有限公司 44102

专利代理师 刘瑶云

(51) Int. Cl.

G06F 16/215 (2019.01)

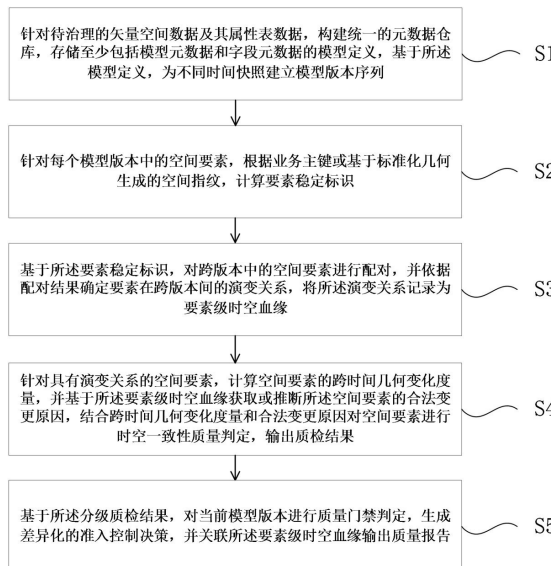
权利要求书3页 说明书17页 附图3页

(54) 发明名称

基于质量控制和血缘追溯的地理空间数据治理方法及系统

(57) 摘要

本发明公开了基于质量控制和血缘追溯的地理空间数据治理方法及系统。方法包括：构建统一元数据仓库，存储模型及字段元数据，并为不同时间快照建立模型版本序列；通过业务主键或基于标准化几何生成的空间指纹，为各版本空间要素计算稳定标识，据此进行跨版本要素配对，识别并记录要素级时空血缘演变关系；进而计算空间要素跨时间几何变化度量，结合合法变更原因的获取或推断，实施时空一致性质量判定并输出分级质检结果；最终依据分级质检结果进行质量门禁判定，生成差异化准入控制决策，并关联要素级时空血缘输出质量报告。本发明实现了地理空间数据在版本演变过程中的自动化血缘追溯与一致性质检，提升了数据治理的可靠性与可审计性。



1. 一种基于质量控制和血缘追溯的地理空间数据治理方法,其特征在于,包括:

针对待治理的矢量空间数据及其属性表数据,构建统一的元数据仓库,存储至少包括模型元数据和字段元数据的模型定义,基于所述模型定义,为不同时间快照建立模型版本序列;

针对每个模型版本中的空间要素,根据业务主键或基于标准化几何生成的空间指纹,计算要素稳定标识;

基于所述要素稳定标识,对跨版本中的空间要素进行配对,并依据配对结果确定要素在跨版本间的演变关系,将所述演变关系记录为要素级时空血缘;

针对具有演变关系的空间要素,计算空间要素的跨时间几何变化度量,并基于所述要素级时空血缘获取或推断所述空间要素的合法变更原因,结合跨时间几何变化度量和合法变更原因对空间要素进行时空一致性质量判定,输出分级质检结果;

基于所述分级质检结果,对当前模型版本进行质量门禁判定,生成差异化的准入控制决策,并关联所述要素级时空血缘以输出质量报告。

2. 根据权利要求1所述的方法,其特征在于,所述模型元数据至少包括空间参考标识、几何类型、拓扑约束集合、版本标识字段以及要素稳定标识生成策略;所述字段元数据至少包含字段约束条件与语义标签,用于规则自动装配与血缘语义标注。

3. 根据权利要求1所述的方法,其特征在于,所述计算要素稳定标识具体包括:

所述要素稳定标识通过以下方式生成:

对原始几何执行网格吸附容差处理与环序规范化处理,得到标准化几何;

对所述标准化几何进行WKB编码并计算哈希值,得到空间指纹;

将业务主键与所述空间指纹拼接后计算哈希值,或在业务主键缺失时直接以所述空间指纹作为要素稳定标识;

计算空间重叠比作为补充配对依据,其中对面要素基于交并面积计算,对线要素与点要素通过缓冲生成带状面或圆域后再计算交并面积得到空间重叠比,或采用线形差异度量或位移距离作为补充配对依据,在所述要素稳定标识无法生成或无法精确匹配时根据所述补充配对依据进行配对。

4. 根据权利要求3所述的方法,其特征在于,所述演变关系根据以下逻辑确定:

当跨版本间存在唯一的要素配对,且所述空间重叠比大于或等于第一阈值时,判定为保持关系;

当跨版本间存在唯一的要素配对,且所述空间重叠比小于所述第一阈值时,判定为更新关系;

当旧版本中的多个要素与新版本中的一个要素配对时,判定为合并关系;

当旧版本中的一个要素与新版本中的多个要素配对时,判定为拆分关系;

当要素仅出现在新版本中时,判定为新增关系;

当要素仅出现在旧版本中时,判定为删除关系。

5. 根据权利要求3所述的方法,其特征在于,所述几何变化度量包括:

对面要素计算面积变化率和对称差异率;

对线要素计算长度变化率和线形差异度量;

对点要素计算位移距离;

所述面积变化率的计算公式为：

$$\Delta A = \frac{|A_{t+1} - A_t|}{A_t}$$

其中, A_t 和 A_{t+1} 分别表示所述空间要素在相邻两个版本中的面积；

所述对称差异率的计算公式为：

$$D = \frac{\text{Area}(\text{SymDiff}(G_t, G_{t+1}))}{\text{Area}(G_t)}$$

其中, G_t 和 G_{t+1} 分别表示所述空间要素在相邻两个版本中的要素几何, SymDiff 表示对称差运算；

所述长度变化率的计算公式为：

$$\Delta L = \frac{|L_{t+1} - L_t|}{L_t}$$

其中, L_t 和 L_{t+1} 分别表示所述空间要素在相邻两个版本中的线要素长度；

所述线形差异度量的计算公式为：

$$D_L = \frac{\text{Area}(\text{SymDiff}(\text{Buffer}(G_t, r), \text{Buffer}(G_{t+1}, r)))}{\text{Area}(\text{Buffer}(G_t, r))}$$

其中, $\text{Buffer}(G_t, r)$ 和 $\text{Buffer}(G_{t+1}, r)$ 分别表示进行半径为 r 的缓冲生成带状面；

所述位移距离的计算公式为：

$$\Delta d = \text{Dist}(P_t, P_{t+1})$$

其中, P_t 和 P_{t+1} 分别表示所述空间要素在相邻两个版本中的坐标位置。

6. 根据权利要求1所述的方法,其特征在于,所述合法变更原因通过以下至少一种方式获取：

从所述空间要素的属性表或关联业务表中提取审批类型字段并映射为标准化的审批变更原因；

从生成所述模型版本的数据处理作业的元数据中,解析出导致几何变化的算子类型并映射为标准化的作业变更原因。

7. 根据权利要求1所述的方法,其特征在于,所述时空一致性质量判定具体包括：

当所述跨时间几何变化度量超过对应阈值且未获取到有效的合法变更原因时,判定为强规则违规；

当所述跨时间几何变化度量超过对应阈值但获取到有效的合法变更原因时,判定为弱规则违规；

以所述合法变更原因为上下文参数,对命中合法变更原因的要素所触发的空间拓扑规则或属性一致性规则采用弱约束模式执行,仅记录预警；

对未获取合法变更原因的要素所触发的规则采用强约束模式执行,直接判定质量异常。

8. 一种基于质量控制和血缘追溯的地理空间数据治理系统,其特征在于,包括：

元数据管理模块,针对待治理的矢量空间数据及其属性表数据,构建统一的元数据仓库,用于存储和管理模型元数据与字段元数据；

版本与标识管理模块,用于基于所述元数据管理模块生成并管理模型版本序列,并为各版本中的空间要素生成要素稳定标识;

时空血缘构建模块,用于基于所述要素稳定标识进行跨版本要素配对,并确定演变关系,形成要素级时空血缘;

时空一致性校验模块,用于针对具有演变关系的空间要素,计算空间要素的跨时间几何变化度量,并基于所述要素级时空血缘获取或推断所述空间要素的合法变更原因,结合跨时间几何变化度量和合法变更原因对空间要素进行时空一致性质量判定,输出分级质检结果;

质量门禁判定与输出模块,用于基于所述分级质检结果,对当前模型版本进行质量门禁判定,生成差异化的准入控制决策,并关联所述要素级时空血缘输出质量报告。

9. 根据权利要求8所述的系统,其特征在于,所述时空一致性校验模块具体用于:

对面要素计算面积变化率和对称差异比;

对线要素计算长度变化率及基于缓冲对称差比的线形差异度量;

对点要素计算位移距离;

当所述跨时间几何变化度量超过对应阈值且未获取到有效的合法变更原因时,判定为强规则违规;

当所述跨时间几何变化度量超过对应阈值但获取到有效的合法变更原因时,判定为弱规则违规。

10. 根据权利要求8所述的系统,其特征在于,所述时空一致性校验模块还用于:

以所述合法变更原因为上下文参数,对命中合法变更原因的要素所触发的空间拓扑规则或属性一致性规则采用弱约束模式执行,仅记录预警;

对未获取合法变更原因的要素所触发的规则采用强约束模式执行,直接判定质量异常。

基于质量控制和血缘追溯的地理空间数据治理方法及系统

技术领域

[0001] 本申请涉及地理空间数据治理技术领域,特别是涉及基于质量控制和血缘追溯的地理空间数据治理方法及系统。

背景技术

[0002] 随着自然资源、国土空间规划、林业、生态环境、房地产等行业业务的信息化与数字化程度提升,地理空间数据在实际应用中呈现出以下特征:一是数据形态多样,既包含点线面矢量要素及其几何属性,又包含大量与空间对象关联的属性表数据;二是数据更新频繁,存在按年度、季度、项目批次等形成的多时间版本数据;三是数据加工链条复杂,包含抽取、融合、空间叠置分析、指标计算等多种处理过程,下游往往依赖上游数据持续更新迭代。

[0003] 现有数据治理平台或质量管理工具通常能够提供属性规则校验(如非空、值域、格式)与部分空间几何检查(如几何有效性、拓扑关系检查),也可能提供表级或字段级的血缘追踪。但在地理空间数据治理中仍普遍存在以下问题:

[0004] (1) 缺少面向多版本数据的时空一致性质量控制。现有质检多以“单次快照”检查为主,难以判断同一空间对象在不同时间版本之间的形变、面积变化、位置偏移等是否合理,无法区分“合法变更”与“异常变更”。

[0005] (2) 血缘追踪粒度不足或缺少时空演变语义。仅有表级或字段级血缘难以回答“某一版本的某条要素从哪些历史要素演变而来”“合并/拆分导致的对象演变链如何追溯”等关键问题,进而导致问题定位依赖人工经验。

[0006] (3) 质量问题与血缘信息割裂,难以形成闭环。质量问题发现后,缺少基于血缘的根因定位、影响范围推导与修复优先级排序机制,导致修复效率低、整改成本高。

[0007] (4) 地理空间数据的专属约束未被模型化沉淀。坐标系、几何类型、Z/M值、拓扑约束、版本轴等空间特性若未在模型层统一定义,往往分散在脚本、配置或人工约定中,难以复用并导致规则执行与血缘建模缺少统一依据。

发明内容

[0008] 基于此,本发明的发明目的是针对上述技术问题,提供基于质量控制和血缘追溯的地理空间数据治理方法及系统,能够将模型定义、时空血缘、跨时间一致性质检、强弱规则分级与报告输出联动起来,实现可追溯、可解释、闭环的地理空间数据治理流程。

[0009] 为实现上述发明目的,本申请第一方面提供一种基于质量控制和血缘追溯的地理空间数据治理方法,包括:

针对待治理的矢量空间数据及其属性表数据,构建统一的元数据仓库,存储至少包括模型元数据和字段元数据的模型定义,基于所述模型定义,为所述矢量空间数据的不同时间快照建立模型版本序列;

针对每个模型版本中的空间要素,根据业务主键或基于标准化几何生成的空间指纹,计算要素稳定标识;

基于所述要素稳定标识,对跨版本中的空间要素进行配对,并依据配对结果确定要素在跨版本间的演变关系,将所述演变关系记录为要素级时空血缘;

针对具有演变关系的空间要素,计算空间要素的跨时间几何变化度量,并基于所述要素级时空血缘获取或推断所述空间要素的合法变更原因,结合跨时间几何变化度量和合法变更原因对空间要素进行时空一致性质量判定,输出分级质检结果;

基于所述分级质检结果,对当前模型版本进行质量门禁判定,生成差异化的准入控制决策,并关联所述要素级时空血缘输出质量报告。

[0010] 优选地,所述模型元数据至少包括空间参考标识、几何类型、拓扑约束集合、版本标识字段以及要素稳定标识生成策略;所述字段元数据至少包含字段约束条件与语义标签,用于规则自动装配与血缘语义标注。

[0011] 优选地,所述计算要素稳定标识具体包括:

所述要素稳定标识通过以下方式生成:

对原始几何执行网格吸附容差处理与环序规范化处理,得到标准化几何;

对所述标准化几何进行WKB编码并计算哈希值,得到空间指纹;

将业务主键与所述空间指纹拼接后计算哈希值,或在业务主键缺失时直接以所述空间指纹作为要素稳定标识;

计算空间重叠比作为补充配对依据,其中对面要素基于交并面积计算,对线要素与点要素通过缓冲生成带状面或圆域后再计算交并面积得到空间重叠比,或采用线形差异度量或位移距离作为补充配对依据,在所述要素稳定标识无法生成或无法精确匹配时根据所述补充配对依据进行配对。

[0012] 优选地,所述演变关系根据以下逻辑确定:

当跨版本间存在唯一的要素配对,且所述空间重叠比大于或等于第一阈值时,判定为保持关系;

当跨版本间存在唯一的要素配对,且所述空间重叠比小于所述第一阈值时,判定为更新关系;

当旧版本中的多个要素与新版本中的一个要素配对时,判定为合并关系;

当旧版本中的一个要素与新版本中的多个要素配对时,判定为拆分关系;

当要素仅出现在新版本中时,判定为新增关系;

当要素仅出现在旧版本中时,判定为删除关系。

[0013] 优选地,所述几何变化度量包括:

对面要素计算面积变化率和对称差异率;

对线要素计算长度变化率和线形差异度量;

对点要素计算位移距离;

所述面积变化率的计算公式为:

$$\Delta A = \frac{|A_{t+1} - A_t|}{A_t}$$

其中, A_t 和 A_{t+1} 分别表示所述空间要素在相邻两个版本中的面积;

所述对称差异率的计算公式为:

$$D = \frac{\text{Area}(\text{SymDiff}(G_t, G_{t+1}))}{\text{Area}(G_t)}$$

[0014] 其中, G_t 和 G_{t+1} 分别表示所述空间要素在相邻两个版本中的要素几何, SymDiff 表示对称差运算;

所述长度变化率的计算公式为:

$$\Delta L = \frac{|L_{t+1} - L_t|}{L_t}$$

[0015] 其中, L_t 和 L_{t+1} 分别表示所述空间要素在相邻两个版本中的线要素长度; 所述线形差异度量的计算公式为:

$$D_L = \frac{\text{Area}(\text{SymDiff}(\text{Buffer}(G_t, r), \text{Buffer}(G_{t+1}, r)))}{\text{Area}(\text{Buffer}(G_t, r))}$$

[0016] 其中, $\text{Buffer}(G_t, r)$ 或 $\text{Buffer}(G_{t+1}, r)$ 分别表示进行半径为 r 的缓冲生成带状面;

所述位移距离的计算公式为:

$$\Delta d = \text{Dist}(P_t, P_{t+1})$$

[0017] 其中, P_t 和 P_{t+1} 分别表示所述空间要素在相邻两个版本中的坐标位置。

[0018] 优选地, 所述合法变更原因通过以下至少一种方式获取:

从所述空间要素的属性表或关联业务表中提取审批类型字段并映射为标准化的审批变更原因;

从生成所述模型版本的数据处理作业的元数据中, 解析出导致几何变化的算子类型并映射为标准化的作业变更原因。

[0019] 优选地, 所述时空一致性质量判定具体包括:

当所述跨时间几何变化度量超过对应阈值且未获取到有效的合法变更原因时, 判定为强规则违规;

当所述跨时间几何变化度量超过对应阈值但获取到有效的合法变更原因时, 判定为弱规则违规;

以所述合法变更原因为上下文参数, 对命中合法变更原因的要素所触发的空间拓扑规则或属性一致性规则采用弱约束模式执行, 仅记录预警;

对未获取合法变更原因的要素所触发的规则采用强约束模式执行, 直接判定质量异常。

[0020] 为实现发明目的, 本申请第二方面提供一种基于质量控制和血缘追溯的地理空间数据治理系统, 应用上述技术方案所述的一种基于质量控制和血缘追溯的地理空间数据治理方法, 所述系统包括:

元数据管理模块, 针对待治理的矢量空间数据及其属性表数据, 构建统一的元数据仓库, 用于存储和管理模型元数据与字段元数据;

版本与标识管理模块, 用于基于所述元数据管理模块生成并管理模型版本序列, 并为各版本中的空间要素生成要素稳定标识;

时空血缘构建模块, 用于基于所述要素稳定标识进行跨版本要素配对, 并确定演

变关系,形成要素级时空血缘;

时空一致性校验模块,用于针对具有演变关系的空间要素,计算空间要素的跨时间几何变化度量,并基于所述要素级时空血缘获取或推断所述空间要素的合法变更原因,结合跨时间几何变化度量和合法变更原因对空间要素进行时空一致性质量判定,输出分级质检结果;

质量门禁判定与输出模块,用于基于所述分级质检结果,对当前模型版本进行质量门禁判定,生成差异化的准入控制决策,并关联所述要素级时空血缘输出质量报告。

[0021] 优选地,所述时空一致性校验模块具体用于:

对面要素计算面积变化率和对称差异比;

对线要素计算长度变化率及基于缓冲对称差比的线形差异度量;

对点要素计算位移距离;

当所述跨时间几何变化度量超过对应阈值且未获取到有效的合法变更原因时,判定为强规则违规;

当所述跨时间几何变化度量超过对应阈值但获取到有效的合法变更原因时,判定为弱规则违规。

[0022] 优选地,以所述合法变更原因为上下文参数,对命中合法变更原因的要素所触发的空间拓扑规则或属性一致性规则采用弱约束模式执行,仅记录预警;

对未获取合法变更原因的要素所触发的规则采用强约束模式执行,直接判定质量异常。

[0023] 本发明与现有技术相比,其有益效果是:

本发明通过模型元数据统一定义空间约束与版本,为跨时间比对提供一致依据;基于要素稳定标识构建要素级时空血缘,精准追溯空间对象在版本间的保持、更新、合并、拆分等演变关系;在时空一致性校验中引入几何变化度量与合法变更原因,生成分级质检结果,有效区分合法调整与异常变更;并将质检结果与血缘信息联动,通过质量门禁机制对模型版本生成差异化准入决策,实现从模型定义、血缘构建、质量判定到数据准入的治理闭环,显著提升多版本地理空间数据一致性校验的准确性、问题追溯的精细度与流程控制的自动化水平,实现可追溯、可解释、闭环的地理空间数据治理流程。

附图说明

[0024] 图1为本发明实施例中一种基于质量控制和血缘追溯的地理空间数据治理方法的流程示意图;

图2为本发明实施例中进行地理空间数据治理并输出质量报告的流程示意图;

图3为本发明实施例中一种基于质量控制和血缘追溯的地理空间数据治理系统的结构示意图。

具体实施方式

[0025] 为了使本申请的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本申请进行进一步详细说明。应当理解,此处描述的具体实施例仅用以解释本申请,并不用于限定本申请。以下实施例用于说明本发明,但不用来限制本发明的范围。

[0026] 实施例1

本申请实施例1提供一种基于质量控制和血缘追溯的地理空间数据治理方法,如图1所示,包括:

S1、针对待治理的矢量空间数据及其属性表数据,构建统一的元数据仓库,存储至少包括模型元数据和字段元数据的模型定义,基于所述模型定义,为不同时间快照建立模型版本序列;

S2、针对每个模型版本中的空间要素,根据业务主键或基于标准化几何生成的空间指纹,计算要素稳定标识;

S3、基于所述要素稳定标识,对跨版本中的空间要素进行配对,并依据配对结果确定要素在跨版本间的演变关系,将所述演变关系记录为要素级时空血缘;

S4、针对具有演变关系的空间要素,计算空间要素的跨时间几何变化度量,并基于所述要素级时空血缘获取或推断所述空间要素的合法变更原因,结合跨时间几何变化度量和合法变更原因对空间要素进行时空一致性质量判定,输出分级质检结果;

S5、基于所述分级质检结果,对当前模型版本进行质量门禁判定,生成差异化的准入控制决策,并关联所述要素级时空血缘输出质量报告。

[0027] 上述步骤S1-S5的具体内容如下:

本发明以统一元数据仓库为核心,围绕“模型定义—版本轴—稳定标识—时空血缘—跨时间一致性校验—质量门禁—报告与追溯”形成闭环治理流程。本发明的地理空间数据治理对象限定为:矢量空间数据(点/线/面)及其属性表数据。

[0028] (一)、定义总体数据结构

首先定义如下数据对象:

模型(Model):平台治理的基本单位,可为实体表或空间—属性统一模型。

[0029] 模型版本(ModelVersion):同一模型在不同时间点形成的数据快照,记为($V_{\{t\}}$)。

[0030] 要素稳定标识(FeatureID):同一空间对象跨版本识别与配对的唯一标识,可由业务主键与空间指纹生成。

[0031] 演变关系(Evolution):描述要素在相邻版本间的状态变化类型,包括保持、更新、合并、拆分、新增、删除。

[0032] 质量规则(QualityRule):用于校验数据质量的规则集合,按严重程度分为强规则与弱规则。

[0033] 质量门禁(QualityGate):基于规则执行结果对数据版本的准入状态进行判定的机制,状态包括通过、预警通过、阻断。

[0034] 构建元数据仓库结构(输入/输出结构之一:模型与字段元数据):

平台建立统一元数据仓库,用于存储模型定义与字段元数据。其关键表结构可包括但不限于:

1) 模型元数据表(ModelMeta)

用于记录模型级空间约束与版本轴定义:

model_id:模型唯一标识。

[0035] model_name:模型名称。

- [0036] srid:空间参考标识。
- [0037] geom_type:几何类型 (POINT/LINESTRING/POLYGON等)。
- [0038] has_z、has_m:Z/M值标识。
- [0039] topo_rule_set:拓扑约束集合标识 (如有效性、自交、重叠、缝隙等规则集合)。
- [0040] version_field:版本标识字段名 (如version_id)。
- [0041] time_field:采集/生效时间字段名 (如collect_time/valid_from)。
- [0042] featureid_policy:FeatureID策略 (业务主键/空间指纹/组合)。
- [0043] security_level、tags:安全等级与标签 ()。
- [0044] 2) 字段元数据表 (FieldMeta)
至少包含:字段英文名称、字段中文名称、字段类型、字段长度、主键、外键、不能为空、单位、默认值、是否键值、字段安全等级、数据字典、字段数据类型、字段数据层级、字段数据格式、字段说明、知识图谱显示、字段标签等。
- [0045] 字段元数据的作用包括:为属性质检规则的自动装配提供依据 (如不能为空、字典约束、格式约束);为血缘图谱中字段节点的语义标注提供依据 (如是否键值、数据类型、层级、标签);为质量报告的分组汇总提供维度 (模型/字段/标签/安全等级)。
- [0046] 构建血缘与质量数据结构 (输入/输出结构之二:血缘与质检结果表):
1) 作业元信息表 (JobMeta)
记录每次抽取或空间运算作业的基本信息:
job_id:作业标识。
- [0047] operator_type:算子类型 (抽取/融合/裁剪/相交/字段计算等)。
- [0048] params_hash:参数摘要 (用于审计与复现)。
- [0049] src_models、tgt_models:输入/输出模型集合。
- [0050] run_time、status:执行时间与状态。
- [0051] 2) 模型级血缘表 (ModelLineage)
用途:记录模型版本之间因作业加工产生的依赖关系,用于回答“某模型版本由哪些上游模型版本加工而来”。
- [0052] src_model_id:上游模型唯一标识 (来源模型ID)。
- [0053] src_version:上游模型版本标识 (来源版本号/快照号/时间戳)。
- [0054] tgt_model_id:下游模型唯一标识 (目标模型ID)。
- [0055] tgt_version:下游模型版本标识 (目标版本号/快照号/时间戳)。
- [0056] job_id:生成该血缘关系的作业标识 (对应作业元信息表中的作业ID)。
- [0057] operator_type:算子类型/作业类型 (如抽取、融合、裁剪、相交、字段计算等)。
- [0058] lineage_time:血缘关系生成时间 (通常为作业完成时间或血缘落库时间)。
- [0059] 3) 字段级血缘表 (FieldLineage)
用途:记录目标字段与源字段之间的一对多映射关系及字段变换逻辑,用于回答“某字段由哪些源字段计算/映射而来”。
- [0060] src_model_id:源模型唯一标识 (来源表/图层所属模型ID)。
- [0061] src_field:源字段英文名称 (来源字段名)。
- [0062] tgt_model_id:目标模型唯一标识 (目标表/图层所属模型ID)。

- [0063] tgt_field:目标字段英文名称(目标字段名)。
- [0064] transform_expr:字段变换表达式/映射规则(例如colB = f2 * 100,或tgt_field = func(src_field))。
- [0065] job_id:生成该字段血缘的作业标识(对应执行该字段计算/映射的作业ID)。
- [0066] 4) 要素级时空血缘表(FeatureLineage)
用途:记录空间要素在跨版本或跨加工过程中的演变关系,用于回答“某版本某条要素从哪些历史要素演变而来”、“是否发生合并/拆分”等。
- [0067] src_feature_id:源要素稳定标识(上游要素FeatureID,可由业务主键/空间指纹生成)。
- [0068] src_version:源要素所属版本标识(来源版本号/快照号/时间戳)。
- [0069] tgt_feature_id:目标要素稳定标识(下游要素FeatureID)。
- [0070] tgt_version:目标要素所属版本标识(目标版本号/快照号/时间戳)。
- [0071] evolve_type:演变类型。
- [0072] KEEP:保持(变化低于阈值,视为同一对象延续)。
- [0073] UPDATE:更新(一对一但发生显著形变或属性变化)。
- [0074] MERGE:合并(多对一)。
- [0075] SPLIT:拆分(一对多)。
- [0076] NEW:新增(仅在新版本出现)。
- [0077] DELETE:删除(仅在旧版本存在)。
- [0078] job_id:产生该演变关系的作业标识(对应作业ID)。
- [0079] operator_type:算子类型/作业类型(用于解释演变来源,如合并、裁剪、融合等)。
- [0080] params_hash:作业参数摘要/哈希(用于审计与复现该演变关系的计算条件)。
- [0081] evolve_time:演变关系生成时间(通常取作业完成时间或血缘落库时间)。
- [0082] 说明:对于合并/拆分的多对多关系,可通过多条记录表达;例如合并时多个src_feature_id对同一个tgt_feature_id建立多条映射记录;拆分时同一个src_feature_id对多个tgt_feature_id建立多条映射记录。
- [0083] 5) 质量规则表(QualityRule)
用途:统一存储质检规则的元信息与执行表达式,支持按规则类型与强弱等级进行调度与报告汇总。
- [0084] rule_id:规则唯一标识。
- [0085] rule_name:规则名称(便于界面展示与报告引用)。
- [0086] rule_type:规则类型。
- [0087] ATTR:属性规则(非空、字典、范围、格式等)。
- [0088] SPATIAL:空间规则(几何有效性、拓扑关系等)。
- [0089] TEMPORAL:时空/跨时间规则(跨版本一致性、变化阈值等)。
- [0090] severity:严重级别。
- [0091] STRONG:强规则(不允许发生,命中即阻断)。
- [0092] WEAK:弱规则(允许发生但需预警提示)。
- [0093] apply_scope:适用范围(按模型/字段/标签等维度绑定规则)。

- [0094] model:绑定到某模型或模型集合。
- [0095] field:绑定到某字段或字段集合。
- [0096] tag:绑定到具有某标签的模型/字段(便于批量规则装配)。
- [0097] rule_expr:规则表达式或脚本引用(SQL片段、空间表达式、脚本路径/标识等)。
- [0098] threshold_set:阈值集合标识(引用阈值配置,用于同规则在不同模型/尺度下差异化阈值)。
- [0099] 6) 质检结果明细表(QualityIssue)
用途:记录每次规则执行产生的违规明细,用于定位具体问题记录/要素及其证据。
- [0100] rule_id:命中的规则标识。
- [0101] severity:命中规则的严重级别(STRONG/WEAK)。
- [0102] rule_type:命中规则的类型(ATTR/SPATIAL/TEMPORAL)。
- [0103] model_id:发生问题的模型标识。
- [0104] version:发生问题的数据版本标识(用于跨时间管理与回溯)。
- [0105] record_id/feature_id:违规对象标识。
- [0106] record_id:表记录主键(用于纯表/属性记录)。
- [0107] feature_id:空间要素稳定标识FeatureID(用于空间要素)。
- [0108] evidence:证据/说明信息。
- [0109] 可包含:变化度量值(如 ΔA 、 D)、命中字段、对比版本、关联 job_id、关联要素对等。
- [0110] check_time:质检执行时间(规则运行时间或结果落库时间)。
- [0111] 7) 质检汇总表(RuleViolationSummary)
用途:用于报告统计与门禁判定的汇总数据,按规则与版本维度聚合违规数量。
- [0112] rule_id:规则标识。
- [0113] severity:规则严重级别(STRONG/WEAK)。
- [0114] model_id:模型标识。
- [0115] version:版本标识。
- [0116] violation_count:违规数量(该规则在该模型版本上的命中次数)。
- [0117] check_time:汇总生成时间(通常与该轮质检批次时间一致)。
- [0118] 构建质量门禁状态机:
为实现“强规则阻断、弱规则预警”的治理闭环,本发明构建了质量门禁状态机(QualityGate),将规则执行结果映射为模型版本的准入状态,以决定该版本数据是否允许进入下游发布、服务或主题计算等环节。与现有仅输出质检结果的方案不同,本发明的门禁机制将“规则配置—执行—判定—输出”联动成闭环,其中强弱规则的定义与分级策略由用户在平台中配置,平台提供默认分级模板但不强制固定。
- [0119] 强弱规则的定义方式(可配置)
本发明中“强规则(STRONG)/弱规则(WEAK)”属于质量规则的严重级别(severity),由用户在规则管理界面或REST API中为每条规则显式设置,或通过规则模板批量继承设置。用户可根据业务容忍度、数据用途、涉密等级、服务发布要求等因素确定严

重级别。

[0120] (1) 配置粒度

强弱规则支持至少以下粒度的用户设定方式：

规则级分级：对每条规则设置severity=STRONG/WEAK；

规则集/模板级分级：对某类规则集（如“拓扑有效性规则集”“属性字典规则集”）统一设定默认严重级别；

模型级覆盖：同一条规则在不同模型上可采用不同严重级别（例如对“核心库模型”为STRONG，对“数据湖模型”为WEAK）；

字段级覆盖（可选）：同一规则在不同字段上可差异化分级（例如主键字段非空为STRONG，备注字段非空为WEAK）；

标签/安全等级驱动分级（可选）：可根据模型或字段标签、数据安全等级自动映射默认严重级别。

[0121] 说明：上述覆盖关系优先级设定为：字段级覆盖>模型级覆盖>规则级设定>规则集/模板默认值。

[0122] (2) 配置入口与存储结构

平台在QualityRule（质量规则表）中存储规则及其严重级别字段severity，并允许用户通过如下方式维护：

图形化界面：规则编辑页中选择“强规则/弱规则”；

API：提交规则元数据时携带severity；

模板导入：从规则模板批量导入并继承默认严重级别。

[0123] 除severity外，规则还可配置以下与门禁相关的参数（用户可选填）：

gate_action：命中后的门禁动作（阻断/预警/仅记录）；

allow_condition：允许条件（例如存在合法变更原因时降级处理）；

threshold_set：阈值集合（允许按模型或比例尺差异化阈值）。

[0124] 强弱规则的常用分级策略（默认建议，可调整）：

为降低用户配置成本，提供默认建议（模板），但最终以用户设定为准。例如：

通常建议为STRONG的规则：几何无效、空几何、主键为空、关键字字典字段超出值域、跨时间异常突变且无合法原因等；

通常建议为WEAK的规则：非关键描述字段为空、轻微几何形变但未超阈值、可接受的缺失字段（业务允许）等。用户可依据不同治理层级设置策略，例如：数据湖层：偏向“先入湖后治理”，部分规则可设为WEAK以提示风险；核心层/发布层：偏向“可用可控”，关键规则设为STRONG以保障下游服务质量。

[0125] 规则命中后的降级/例外处理（用户可配置）：

为避免“合理变更被误判为阻断”，门禁逻辑支持用户定义规则的降级策略：

当severity=STRONG的规则命中时，若满足allow_condition（例如具备合法变更原因、审批通过、属于指定算子类型、在白名单时段/区域内等），则可将该次命中从STRONG降级为WEAK，或仅记录不阻断；对跨时间一致性规则（TEMPORAL类），可配置“合法变更原因”字段或判定条件，作为门禁判定的重要证据链。

[0126] 质量门禁状态机与判定逻辑：

门禁状态用于对“模型版本”给出准入结论,至少包括以下三种状态:PASS(通过):未命中任何STRONG或WEAK;WARN(预警通过):未命中STRONG,但命中WEAK;FAIL(阻断):命中至少一条STRONG(或命中STRONG次数超过阈值)。

[0127] 其判定逻辑示例为:

$$GateStatus(V_t) = \begin{cases} FAIL, & StrongCount(V_t) > 0 \\ WARN, & StrongCount(V_t) = 0 \wedge WeakCount(V_t) > 0 \\ PASS, & StrongCount(V_t) = 0 \wedge WeakCount(V_t) = 0 \end{cases}$$

[0128]

参数/符号	含义	备注/典型取值
GateStatus(V _t)	质量门禁判定结果函数 (对版本(V _t)的准入状态)	输出为 FAIL/WARN/PASS
V _t	某模型在时间点/版本(t)的数据快照(待判定的数据版本)	(t)可为时间戳/批次号/版本号
StrongCount(V _t)	版本(V _t)中强规则 (STRONG)命中次数 (违规数)	一般按“命中事件数/违规记录数/违规要素数”统计
(WeakCount(V _t))	版本(V _t)中弱规则 (WEAK)命中次数(违规数)	同上,统计口径与 StrongCount 保持一致
FAIL	阻断(不允许进入下游流程)	通常表示强规则命中(或超阈值)
WARN	预警通过(允许进入下游,但提示风险)	常见条件:强规则为0且弱规则>0
PASS	通过(允许进入下游且无预警项)	常见条件:强规则为0且弱规则为0

[0129] 表1 质量门禁状态机的判定逻辑示例的参数定义表

表1展示了质量门禁状态机的判定逻辑示例的参数定义,其中StrongCount(V_t)与WeakCount(V_t)来自对QualityIssue或RuleViolationSummary的聚合统计;聚合维度至少包括model_id与version。在启用降级策略时,统计应以“降级后的最终严重级别”为准。

[0130] 门禁动作与下游联动:

当门禁状态生成后,系统执行与状态对应的动作(可配置):FAIL:阻断该模型版本进入下游发布/服务/主题计算;可同时自动生成修复清单与工单;WARN:允许进入下游,但

将预警规则与典型问题写入报告,并在数据服务侧返回质量提示;PASS:正常流转并记录批次通过信息。上述门禁结果与规则命中明细、血缘链路(model/field/feature lineage)形成关联,支撑后续“根因回溯、影响分析与修复优先级排序”,从而实现治理闭环。

[0131] (二)、基于以上数据结构的具体实施步骤,如图2所示,包括:

步骤1:模型定义与版本轴构建:

通过图形化界面或API创建模型定义,写入ModelMeta与FieldMeta。依据version_field与time_field将同一模型的不同快照组织为版本序列($V\{t_1\}, V\{t_2\}, \dots, V\{t_n\}$)。版本可按自然时间(天/月/年)或按批次号组织。

[0132] 步骤2:FeatureID生成与跨版本配对:

对每个版本中的空间要素生成FeatureID:

(1)几何标准化:

$G_{norm} = \text{Normalize}(\text{SnapToGrid}(G, \epsilon), \text{OrderRing}(G))$

[0133] 在几何标准化公式中, G 表示输入的原始几何对象(点/线/面及其坐标序列); ϵ 表示网格吸附容差(将坐标按给定精度进行量化/舍入,使相近坐标归并到同一网格以消除浮点抖动与微小噪声); $\text{SnapToGrid}(G, \epsilon)$ 表示对 G 的所有顶点坐标执行以 ϵ 为分辨率的网格化吸附处理; $\text{OrderRing}(G)$ 表示对面要素的环(外环与内环)及其顶点顺序进行规范化排序(例如统一外环方向、内环方向,并按稳定规则选择起始顶点以消除“同形不同序”的表示差异); $\text{Normalize}()$ 表示将吸附后的几何与环序规范结果组合,输出满足坐标精度与拓扑表示一致性的标准化几何 G_{norm} ,用于后续空间指纹/哈希与跨版本要素配对。

[0134] (2)空间指纹:

$fp = \text{Hash}(\text{WKB}(G_{norm}))$

[0135] 在空间指纹公式 $fp = \text{Hash}(\text{WKB}(G_{norm}))$ 中, G_{norm} 表示经过几何标准化后的几何对象(已完成网格吸附、环序/方向统一等规范化处理,用于消除同形不同序与浮点误差影响); $\text{WKB}()$ 表示将几何对象按Well-Known Binary标准编码为稳定的二进制字节序列,以便在不同系统/存储引擎间获得一致的几何表示; $\text{Hash}()$ 表示对该二进制字节序列执行哈希摘要计算(例如MD5、SHA-256或其他可配置哈希算法),输出固定长度的摘要值作为空间指纹; fp 即空间指纹值,用于在跨版本配对、去重、快速比对与血缘关联中作为几何形态的稳定标识。

[0136] (3)FeatureID:

$\text{FeatureID} = \text{Hash}(\text{biz_id}\|fp)$ 或 $\text{FeatureID} = fp$

[0137] 在要素稳定标识公式($\text{FeatureID}=\text{Hash}(\text{biz_id}\|fp)$)或($\text{FeatureID}=fp$)中, (FeatureID)表示用于跨版本配对与时空血缘追溯的要素唯一稳定标识; (biz_id)表示业务主键或外部系统赋予的对象唯一编码(例如不随版本变化的地块/宗地/对象ID),用于提供业务语义上的稳定性; (fp)表示由标准化几何 G_{norm} 生成的空间指纹摘要,用于提供几何形态上的稳定性;符号 $\|$ 表示字符串/字节序列拼接(通常采用固定分隔符或长度前缀以避免歧义),即将(biz_id)与(fp)组合形成联合特征; $\text{Hash}()$ 为哈希摘要算法,用于将联合特征映射为固定长度的唯一标识值,实际实现中可选用MD5、SHA-256或MurmurHash等(可配置),其中面向安全性与抗碰撞要求较高的场景优先采用SHA-256,面向高性能大规模去重与配对的场景可采用MurmurHash;当数据缺少可靠业务主键或需仅以几何形态驱动配对

时,可直接令 $\text{FeatureID} = \text{fp}$,以空间指纹作为要素标识,但此时需结合版本轴与必要的辅助条件(如属性键值、空间邻接关系)降低哈希碰撞或同形异物导致的误配风险。

[0138] 跨版本配对优先使用业务主键或FeatureID精确匹配;若缺失,则采用空间重叠比匹配:

$$R = \frac{\text{Area}(\text{Intersection}(G_t, G_{t+1}))}{\text{Area}(\text{Union}(G_t, G_{t+1}))}$$

[0139] 计算空间重叠比作为补充配对依据,其中对面要素基于交并面积计算,对线要素与点要素通过缓冲生成带状面或圆域后再计算交并面积得到空间重叠比,或采用线形差异度量或位移距离作为补充配对依据,在要素稳定标识无法生成或无法精确匹配时根据所述补充配对依据进行配对。线/点要素可先按半径 r 缓冲为面后计算 R , r 由采集精度或值集配置。在跨版本配对过程中,平台对相邻版本 V_t 与 V_{t+1} 的要素首先采用确定性精确匹配策略:若存在稳定的业务主键 biz_id 或已生成的要素稳定标识FeatureID,则以“主键相等/FeatureID相等”直接建立一对一候选对应关系;当上述键缺失、质量不可靠或出现一键多要素等异常时,转入基于空间相似度的候选匹配,计算空间重叠比 R 作为匹配度指标,其中 G_t 与 G_{t+1} 分别表示同一模型在版本 t 与 $t+1$ 中的候选要素几何(通常使用标准化后的几何 G_{norm} 以消除微小坐标抖动影响), $\text{Intersection}(G_t, G_{t+1})$ 表示两几何的相交区域, $\text{Union}(G_t, G_{t+1})$ 表示两几何的并集区域, $\text{Area}()$ 表示面积度量运算;该指标本质上等价于Jaccard相似系数(IoU, Intersection over Union),取值范围为 $0 \sim 1$,当 R 越接近1表示空间形态越一致。平台可设置阈值 θR (由用户配置、可按模型类型/尺度分级),当 $R \geq \theta R$ 时将该对要素判定为同一对象的候选配对;当存在多个候选满足阈值时,可进一步采用“最大RRR”优选策略,必要时叠加辅助约束(如同一行政区编码、相邻关系一致、属性键值相等等)以降低误配,并为后续“合并/拆分”演变识别提供候选集合输入。

[0140] 步骤3:构建要素级时空血缘:

对相邻版本对 $((V_t, V_{t+1}))$,依据配对关系与空间相似度(这里的空间相似度指的就是上文定义的空间重叠比指标 R (即基于Intersection/Union的Jaccard相似系数),对应的判定阈值即为 θR)判定演变类型:一对一且变化低于阈值:KEEP。一对一且变化超过阈值:UPDATE。多对一:MERGE。一对多:SPLIT。仅新版本存在:NEW。仅旧版本存在:DELETE。并将映射写入FeatureLineage,同时写入JobMeta、ModelLineage、FieldLineage以形成审计链。

[0141] 步骤4:跨时间一致性校验(TEMPORAL规则):

对相邻版本对 $((V_t, V_{t+1}))$ 中已完成跨版本配对的要素,系统在要素级时空血缘的基础上执行跨时间一致性校验。为适配矢量数据点/线/面不同几何特性,本步骤对跨时间变化度量采用按几何类型自适应的策略:对面要素采用面积类与面域差异类度量;对线要素采用长度类与线形差异类度量;对点要素采用位置位移类度量,并分别与对应阈值比较以识别异常突变与合理变更。

[0142] 对配对要素计算跨时间变化度量:

(1) 针对面要素的跨时间几何变化度量:

面积变化率:

$$\Delta A = \frac{|A_{t+1} - A_t|}{A_t}$$

[0143] 在相邻版本要素演变判定中,引入面积变化率指标 ΔA 用于量化同一配对要素在版本 V_t 与 V_{t+1} 之间的几何尺度变化程度。该指标以版本 t 的要素面积作为归一化基准,计算两期面积的相对变化比例,从而消除绝对面积差异对不同尺度要素的影响。算法上,首先依据业务主键、FeatureID或空间重叠比 R 建立跨版本一对一配对关系,其次分别计算配对要素在 t 与 $t+1$ 时刻的几何面积 A_t 与 A_{t+1} ,并据此得到面积变化率 ΔA 。当 $\Delta A \leq \theta A$ (面积变化阈值)时,可判定为形态稳定或轻微调整;当 $\Delta A > \theta A$ 时,则判定为显著几何变化,通常与边界修正、裁剪扩展或规划调整等演变类型相关。该指标与空间重叠比 R 联合使用,用于区分“高重叠但面积变化显著”与“低重叠导致的结构性变化”两类情形。

[0144] 对称差异率:

$$D = \frac{\text{Area}(\text{SymDiff}(G_t, G_{t+1}))}{\text{Area}(G_t)}$$

[0145] 在跨版本要素演变判定过程中,为进一步刻画几何形态变化的“空间偏移与重构程度”,本发明在面积变化率与空间重叠比的基础上,构建对称差异率 D 。该指标通过计算相邻版本要素几何 G_t 与 G_{t+1} 的对称差(Symmetric Difference)面积,并以基准版本要素面积进行归一化,用于量化“非重叠部分”在原要素中的占比。对称差运算能够同时捕捉新增与删减的几何区域,避免仅依赖交并比对局部变化不敏感的问题。当 D 较小且低于用户设定的阈值 θD 时,可判定为轻微几何调整(如边界微调、坐标修正);当 D 显著增大并超过阈值时,则表明要素发生了实质性形态重构,作为UPDATE、SPLIT或MERGE等演变类型的重要判据之一。该算法适用于面状要素,基于标准GIS拓扑算子实现,计算过程可在数据库空间函数或GIS引擎中高效完成。

[0146] (2) 针对线要素的跨时间几何变化度量:

对线要素(LINESTRING),系统以版本 t 的线要素长度 L_t 为基准,计算跨时间长度变化率(ΔL),用于刻画道路/河流/管线等线状对象在相邻版本间的长度增减程度。当($\Delta L \leq \theta L$)时判定为线长度变化可接受;当($\Delta L > \theta L$)时判定为显著变化。为进一步刻画线形偏移或几何重构(即长度相近但走向变化较大),系统可选用“缓冲对称差比”作为线形差异度量:对(G_t)与(G_{t+1})分别进行半径为 r 的缓冲生成带状面 $\text{Buffer}(G_t, r)$ 、 $\text{Buffer}(G_{t+1}, r)$,再计算对称差面积占比(D_L)。其中缓冲半径(r)可由数据采集精度、比例尺或模型阈值集配置确定,用于将线形差异转化为可度量的面域差异。该机制适用于线要素在边线平移、弯曲修正、断接调整等情况下的跨时间一致性判定。

$$[0147] \quad \Delta L = \frac{|L_{t+1} - L_t|}{L_t}$$

$$[0148] \quad D_L = \frac{\text{Area}(\text{SymDiff}(\text{Buffer}(G_t, r), \text{Buffer}(G_{t+1}, r)))}{\text{Area}(\text{Buffer}(G_t, r))}$$

[0149] (三) 针对点要素的跨时间几何变化度量:

对点要素(POINT),系统计算相邻版本间的位移距离(Δd)作为跨时间一致性度量,其中(P_t)与(P_{t+1})分别为同一配对点在版本(t)与($t+1$)的坐标位置。距离计算可采用平面距离或椭球测地距离(由模型SRID与度量要求确定)。当($\Delta d \leq \theta d$)时判定为位置变化可接受;当($\Delta d > \theta d$)时判定为显著位移变化,进入后续规则分级与合法原因解释。该机制适用于POI点、监测点、地籍界址点等对象在跨版本更新时的异常迁移识别。

[0150] $\Delta d = \text{Dist}(P_t, P_{t+1})$

[0151] 合法变更原因change_reason的引入与解释性判定:

在完成相邻版本要素 (G_t) 与 (G_{t+1}) 的几何变化度量后,系统将度量结果与阈值比较输出为TEMPORAL规则命中结果,并引入合法变更原因标识change_reason作为时间一致性判定的修正因子。change_reason并非人工随意赋值,而是通过以下两类途径自动引入:

(1) 审批/业务字段映射:从要素属性表或关联业务表中提取与该要素绑定的审批类型字段(如“调整类型”“批复事项”“变更依据”等),并映射为标准化的变更原因编码;

(2) JobMeta作业元数据解析:从数据生产或治理流程中生成的JobMeta元数据中,解析当前要素所经历的算子链路(如“边界校正”“裁剪”“缓冲”“拓扑修复”“版本替换”等),并根据算子类型规则推断其几何变化性质。

[0152] 当提取到的change_reason位于允许列表中时,即便面/线/点对应的变化度量 ($\Delta A, D, \Delta L, D_L, \Delta d$) 超过阈值,也可将该变化标记为可解释变化,在时间一致性层面予以放行或降级处理;若无法提取合法原因,或原因不在允许范围内,则仍严格依据阈值结果判定为显著变化,并将证据(变化度量值、对比版本、要素对等)写入QualityIssue,为后续空间与属性规则提供输入。

[0153] 步骤5:空间与属性规则执行 (SPATIAL/ATTR规则):

在TEMPORAL规则输出的基础上(面/线/点分别对应其跨时间变化度量与阈值判定结果),系统进一步联合空间拓扑规则与属性一致性规则,对变化要素进行精细化判定。空间规则侧重点在于检验几何变化是否引发拓扑冲突(如自相交、缝隙、重叠、越界等),以及是否符合目标图层的空间约束要求;属性规则则用于校验关键属性(如权属编码、用途类型、来源标识等)在变更前后的逻辑一致性。

[0154] 在该阶段,change_reason被作为规则执行的上下文参数参与判定流程:若其来源于审批字段或JobMeta算子推理,且属于允许列表中的类型,则部分空间或属性规则可采用弱约束模式(WEAK)执行,例如仅记录警告而不直接阻断;反之,则采用强约束模式(STRONG),对不符合规则的要素直接判定为质量异常。通过该机制,实现“规则强度随变更语义动态调整”,避免一刀切式的误判。

[0155] 步骤6:质量门禁判定与报告输出:

在综合TEMPORAL、SPATIAL与ATTR规则结果后,系统进入质量门禁判定阶段。该阶段按照“规则触发结果+变更原因合法性”双维度进行决策:(1)当几何变化指标超阈值,且change_reason不存在或不在允许列表中时,判定为STRONG异常,直接阻断数据入库或版本发布;(2)当几何变化指标超阈值,但change_reason合法且可解释时,判定为WEAK异常,允许入库但需在质检报告中留痕;(3)当几何变化指标未超阈值时,无论是否存在change_reason,均判定为通过。最终,系统自动生成结构化质检报告,明确记录要素标识、变化指标值、触发规则、change_reason来源(审批字段或JobMeta算子)、判定等级(STRONG / WEAK / PASS)及处理建议,其中变化指标值随几何类型输出:面要素输出 $\Delta A, D$ (及必要时R);线要素输出 $\Delta L, D_L$;点要素输出 Δd 。实现全过程可追溯、可审计、可复核的质量门禁控制。

[0156] 实施例2

本申请实施例2基于实施例1,提供一种基于质量控制和血缘追溯的地理空间数据

治理系统,如图3所示,包括:

元数据管理模块101,用于针对待治理的矢量空间数据及其属性表数据,构建统一的元数据仓库,用于存储和管理模型元数据与字段元数据;

版本与标识管理模块201,用于基于所述元数据管理模块生成并管理模型版本序列,并为各版本中的空间要素生成要素稳定标识;

时空血缘构建模块301,用于基于所述要素稳定标识进行跨版本要素配对,并确定演变关系,形成要素级时空血缘;

时空一致性校验模块401,用于针对具有演变关系的空间要素,计算空间要素的跨时间几何变化度量,并基于所述要素级时空血缘获取或推断所述空间要素的合法变更原因,结合跨时间几何变化度量和合法变更原因对空间要素进行时空一致性质量判定,输出分级质检结果;

质量门禁判定与输出模块501,用于基于所述分级质检结果,对当前模型版本进行质量门禁判定,生成差异化的准入控制决策,并关联所述要素级时空血缘输出质量报告。

[0157] 所述时空一致性校验模块401具体用于:对面要素计算面积变化率和对称差异比;对线要素计算长度变化率及基于缓冲对称差比的线形差异度量;对点要素计算位移距离;当所述跨时间几何变化度量超过对应阈值且未获取到有效的合法变更原因时,判定为强规则违规;当所述跨时间几何变化度量超过对应阈值但获取到有效的合法变更原因时,判定为弱规则违规;以所述合法变更原因为上下文参数,对命中合法变更原因的要素所触发的空间拓扑规则或属性一致性规则采用弱约束模式执行,仅记录预警;对未获取合法变更原因的要素所触发的规则采用强约束模式执行,直接判定质量异常。

[0158] 实施例3

本实施例3基于实施例1,以某省自然资源主管部门统一管理的县级行政区划边界数据治理为例对本发明方法进行进一步地说明,具体如下:

该行政区划边界数据以面要素形式存储,按年度形成多个版本,用于国土空间规划、用途管制、统计分析及下游专题应用。由于行政区划调整、边界勘界修正、数据精度提升等原因,不同年度版本之间可能发生边界微调、行政单元合并或拆分等变化。

[0159] 该类数据具有以下典型特征:面要素数量较多,空间范围连续,拓扑关系敏感;多数要素具备稳定业务主键(如行政区划代码),但在历史数据中存在缺失或变更情况;合并、拆分与边界调整往来源于正式审批文件,但审批信息与空间数据并非始终强关联;下游系统对边界数据的时序一致性与准确性要求较高。因此,适合采用本发明所述方法进行跨时间一致性校验与血缘追溯。

[0160] (1) 模型与版本轴构建:

在GIS平台中定义“行政区划边界”模型,对应 ModelMeta 表中记录如下关键信息:

geom_type =POLYGON:要素几何类型,表示该模型所管理的数据为面状空间要素。

[0161] srid = CGCS2000:空间参考标识,表示该模型采用的坐标参考系统为2000国家大地坐标系(CGCS2000)。

[0162] topo_rule_set={几何有效性、不自交、无缝隙、不重叠}:拓扑规则集合,用于约束面要素的空间拓扑关系,包括:几何形态合法、单要素不发生自相交、相邻要素之间不产生

缝隙、同层要素之间不得相互重叠。

[0163] `version_field=version_id`:版本标识字段,用于区分同一模型在不同时间或不同批次下的数据版本。

[0164] `time_field=effective_date`:时间属性字段,用于描述数据版本或要素的生效时间或时点,作为时序分析与跨时间一致性校验的依据。

[0165] `featureid_policy=biz_id+空间指纹`:要素唯一标识生成策略,通过业务主键标识与空间几何指纹的组合生成稳定的要素标识,用于跨版本要素匹配与血缘追踪。

[0166] 将2022年、2023年两个年度的数据分别作为模型版本(V{2022})与(V{2023}),形成相邻版本对用于后续比对。

[0167] (2) FeatureID生成与跨版本配对:

对(V{2022})与(V{2023})中的所有面要素,系统首先执行几何标准化处理,包括坐标网格吸附与环序规范化,得到标准化几何(G_{norm}),并基于其WKB表示生成空间指纹fp。

[0168] 随后,按照以下策略生成要素稳定标识FeatureID:

对存在稳定行政区划代码的要素,采用 $FeatureID = Hash(biz_id||fp)$;

对缺失或不可信业务主键的要素,采用 $FeatureID = fp$ 。

[0169] 在跨版本配对阶段,系统优先基于FeatureID执行精确匹配;对未匹配成功的要素,则计算空间重叠比(R),当 $R \geq \theta_R$ 时建立候选配对关系,并在多候选情况下选择重叠比最大的匹配对。

[0170] (3) 要素级时空血缘构建:

基于跨版本配对结果,系统对相邻版本对((V{2022}, V{2023}))构建要素级时空血缘关系:对一对一匹配且空间相似度高的要素,判定为KEEP或UPDATE;对多个(V{2022})要素对应一个(V{2023})要素的情况,判定为MERGE;对一个(V{2022})要素对应多个(V{2023})要素的情况,判定为SPLIT;仅在新版本出现的要素,标记为NEW;仅在旧版本存在的要素,标记为DELETE。所有演变关系写入FeatureLineage表,并关联对应的job_id与operator_type,以形成可审计的演变链路。

[0171] (4) 跨时间一致性校验与合法变更识别:

针对一对一演变关系的要素,系统计算面积变化率(ΔA)与对称差异率(D),量化边界变化程度。当变化指标超过配置阈值时,系统进一步引入合法变更原因change_reason进行解释性判定。具体包括:系统从行政区划属性表中读取“调整类型”“批复文号”等字段,映射为标准化变更原因编码;同时解析JobMeta中记录的算子类型,例如“行政区合并”“边界修正”“历史数据替换”等,并据此推断变化来源;若提取的change_reason属于允许列表(如“行政区划调整批复”“勘界成果更新”),则认定该变化具备业务合理性。

[0172] (5) 空间与属性规则联合判定:

在TEMPORAL规则结果基础上,系统继续执行空间拓扑规则与属性一致性规则:对因合法变更导致的几何变化,空间规则采用WEAK模式,仅记录预警;对无合法原因支撑的显著变化,空间规则与属性规则均采用STRONG模式;对行政区划代码、名称等关键属性发生不一致且无审批依据的情况,直接判定为质量异常。

[0173] (6) 质量门禁判定与结果输出:系统汇总所有规则执行结果,对(V{2023})版本执

行质量门禁判定:若存在STRONG规则命中,门禁状态为FAIL,阻断该版本进入发布环节;若仅命中WEAK规则,门禁状态为WARN,允许发布但生成预警报告;若未命中任何规则,门禁状态为PASS。

[0174] 同时,系统输出包含以下内容的质量与血缘报告:

要素级演变类型(KEEP/UPDATE/MERGE/SPLIT等);跨时间变化指标(ΔA)、(D)、(R));触发规则及其严重级别;合法变更原因来源与解释路径;关联的历史版本要素及下游影响范围。

[0175] 通过上述实施例可以看出,本发明能够在真实国土空间数据场景下,实现行政区划边界数据的跨时间一致性校验、可解释演变判定与质量门禁控制,有效提升多版本空间数据治理的准确性、可追溯性与自动化水平。

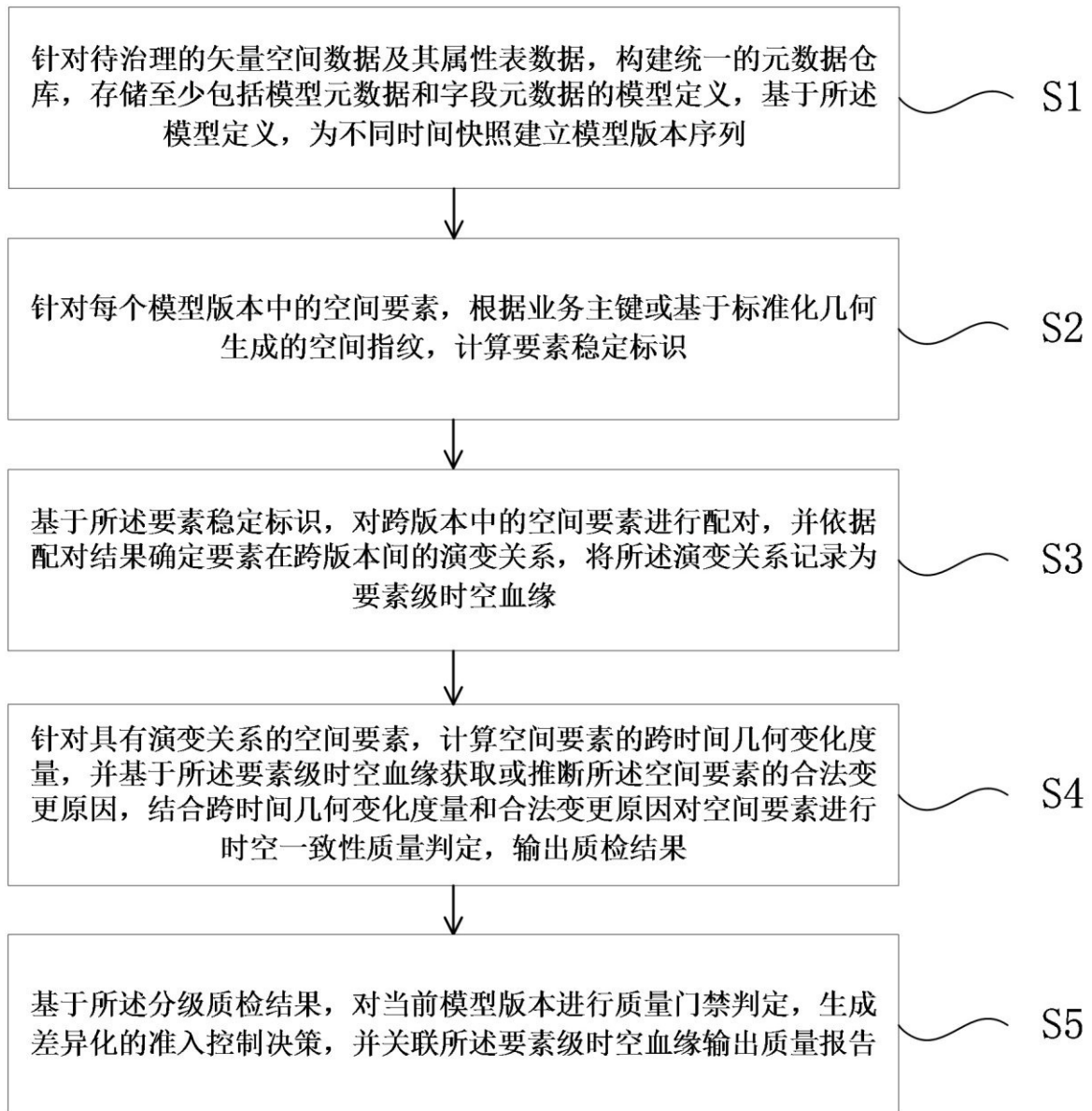


图1

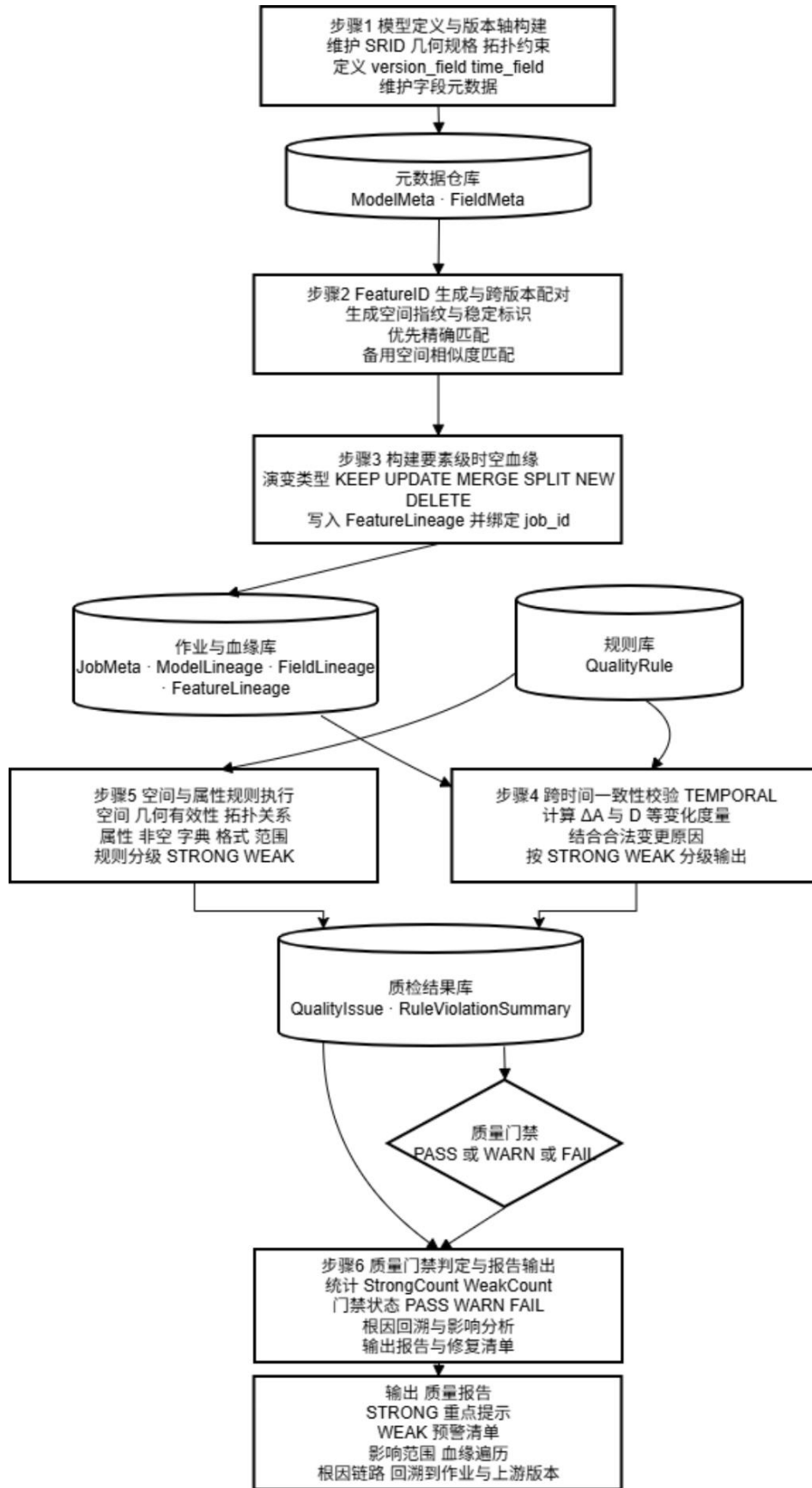


图2

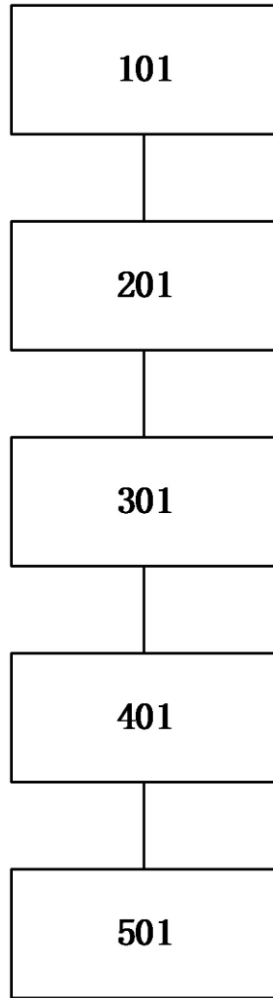


图3