

基于 tableau 结点封闭值的非一致性数据库开放分支修复方法

高 龙¹ 刘 全^{1,2} 傅启明¹ 李 娇¹

(苏州大学计算机科学与技术学院 苏州 215006)¹

(吉林大学符号计算与知识工程教育部重点实验室 长春 130012)²

摘 要 在将 tableau 方法扩展到非一致性数据库修复的基础上,提出一种新的利用分支封闭值修复数据库的方法。该方法结合 tableau 分析法的开放和封闭推理标准,以开放公式树 $TP(IC \cup r)$ 分支为基础,为公式树 $TP(IC \cup r)$ 中每个结点引入一个结点封闭值。根据 $TP(IC \cup r)$ 中结点封闭值的定义,通过计算 $TP(IC \cup r)$ 的结点封闭值来选择分支进行开放修复,从而可以直接确定数据库的修复实例,同时考虑了含有 I 封闭的修复,将开放修复扩展到含有 I 封闭的 $TP(IC \cup r)$,并给出逻辑证明。最后,对于一致性应答结果的逻辑特征予以证明。

关键词 tableau,非一致性数据库,封闭值,开放,修复

中图法分类号 TP311.13 **文献标识码** A

Method of Inconsistent Database Opening Branches Repairs Based on tableau Node Closed Value

GAO Long¹ LIU Quan^{1,2} FU Qi-ming¹ LI Jiao¹

(School of Computer Science and Technology, Soochow University, Suzhou 215006, China)¹

(Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China)²

Abstract Based on extending the tableau method to inconsistent database repairs, a new method that uses the value of branch closure to repair the database was proposed. This method combines the analysis of tableau open and closed reasoning criterion, bases on opening formula tree $TP(IC \cup r)$, and imports a closed value for every node of the $TP(IC \cup r)$. According to the definition of node closed value, branches repair can be selected by computing node closed value, and directly determine the database instances which need repair. This method also considers repair with I closure and extends opening branches to repair, and gives the logical proof. At last, the logical feature of consistent query answer is proved.

Keywords tableau, Inconsistent database, Closed value, Open, Repair

1 引言

在每一个数据库应用中,完整性约束(Integrity Constraint, IC)^[1-3]都是一个极其重要的标准。完整性约束有效地保证了数据的一致性。但是,由于非一致性数据库(Inconsistent Database, IDB)实例的存在,使得满足完整性约束 IC 的数据库实例得不到保证^[1],很多情况会导致数据库处于非一致的状态,如数据从多个独立的数据源被集成到一个统一的数据集合中,会产生语法与语义的不一致;对遗留的数据库系统添加新的完整性约束时,可能存在违反约束的数据;在数据库更新时出现异常也会带来数据的不一致等问题。非一致性的数据库如果用于查询与分析,可能得到错误的结果,直接影响企业业务处理。在这种情况下,获取满足完整性约束的一致性查询应答(Consistent Query Answer, CQA)^[1-3]是非常

重要的。非一致性数据库问题正是数据管理领域的一个挑战。目前主要有两种方法来解决非一致性数据库问题,即数据清洗和一致性查询。

1)数据清洗:首先确定数据库中不一致、错误的属性,然后对数据进行修改,使其恢复到一致状态。然而数据清洗具有一定的局限性。不一致的原因多种多样,对数据的修复需要大量的额外信息,而不同的外界信息往往导致不同的修复结果,因而需要具有丰富的业务知识的人员来参与,使得无法完全自动化地进行数据清理;同时,某些数据清洗操作需要对原数据进行修改,因而不适用于具有重要价值的只读原始数据。

2)一致性查询:在不修改原数据库的情况下进行查询得到满足完整性约束的结果^[4-8]。一致性查询假设原数据库存在多个虚拟(virtual)的修复,查询的结果为这些虚拟的修复

到稿日期:2012-06-11 返修日期:2012-07-17 本文受国家自然科学基金项目(61070223, 61103045, 61070122, 61272005),江苏省自然科学基金(BK2012616),江苏省高校自然科学研究项目(09KJA520002, 09KJB520012),吉林大学符号计算与知识工程教育部重点实验室项目(93K172012K04)资助。

高 龙(1988—),男,硕士,主要研究方向为自动推理、模式识别, E-mail: longgao_88@163.com; 刘 全(1969—),男,博士后,教授,博士生导师, CCF 高级会员,主要研究方向为智能信息处理、自动推理、机器学习; 傅启明(1985—),男,博士, CCF 学生会会员,主要研究方向为自动推理、机器学习; 李 娇(1985—),女,硕士,主要研究方向为自动推理、数据修复。

$It_1) \wedge Product(It_1, T_4) \rightarrow C=C$ 和 $(Supply(C, D_2, It_2) \wedge Product(It_2, T_4) \rightarrow D=C)$ 是通过 γ -规则获取的, 对 $(Supply(C, D_1, It_1) \wedge Product(It_1, T_4) \rightarrow C=C)$ 应用 β -规则产生 3 个子式 $\neg Supply(C, D_1, It_1)$, $\neg Product(It_1, T_4)$ 和 $C=C$ 。整个 $TP(IC \cup r)$ 封闭。

在数据库理论中, 我们做如下假设: (a) 唯一命名假设 (Unique Names Assumption, UNA); 如果 a 与 b 是域 D 中不同的常量, 则在 r 中 $a \neq b$ 。(b) 封闭世界假设 (Closed World Assumption, CWA); 如果 r 是一个数据库实例, 那么对于任意的数据库基原子 $P(c)$, 如果 $P(c) \notin r$, 则 $\neg P(c) \in r$, 更准确地说, 隐式存在 $\neg P(c) \in r$ 。

定义 6 B 为一个带有完整性约束 IC 的数据库实例 r 的分支, 即 $B=IC \cup r$ 。如果 B 是封闭的, 当且仅当满足下列条件之一:

- 1) $a=b \in B$, a 和 b 是 D 中不同的常量。
- 2) (a) $P(\bar{c}) \in I$ 且 $P(\bar{c}) \notin r$, 基元组 $\bar{c}$ 仅由 D 中的元素组成。
- (b) $P(\bar{c}) \in I$ 且在 $\bar{c}$ 中不存在替换 σ 使得 $P(\bar{c}) \in I$ 。
- 3) $\neg P(\bar{c}) \in I$ 且 $P(\bar{c}) \in r$, 基元组 $\bar{c}$ 仅由 D 中的元素组成。
- 4) $\varphi \in B$ 且 $\neg \varphi \in B$, 其中 φ 为任意的公式。
- 5) $\neg t=t \in B$, 其中 t 为任何项。

条件 1) 考虑 UNA 假设, 条件 2) 考虑的是 CWA 假设。对于数据库实例 r 和完整性约束 IC , 如果 r 是关于 IC 非一致的, 当且仅当 $TP(IC \cup r)$ 是封闭的; 如果 $TP(IC \cup r)$ 是封闭的, 当且仅当 r 不满足 IC 。

4 封闭值开放 $TP(IC \cup r)$

一个 tableau 分支是封闭的, 如果它包含一个公式和它的否定, 否则它是开放的。对于数据库实例 r 和完整性约束 IC , 如果非一致性数据库 $r \neq IC$, 则 $TP(IC \cup r)$ 的所有分支封闭。

4.1 数据封闭分支开放

公式树 $TP(IC \cup r)$ 仅用它的封闭分支突出了非一致数据库 r 的特性。为了获得数据库 r 的一个修复, 可以删除分支上导致数据库不一致的文字, 甚至是满足 CWA 假设的隐含文字。根据这种方式, 每一个被开放的分支都将可能产生一个修复。因为严格的完整性约束是只允许改变数据库实例, 所以只能修复由 r 中文字所产生的非一致数据, 而不能删除 I 中的文字。

定义 7 数据库实例 r 是非一致的, 当且仅当 $TP(IC \cup r)$ 是封闭的; $TP(IC \cup r)$ 是封闭的, 当且仅当 r 不满足 $IC(r) \neq IC$ 。

定理 1 数据库实例 r 和公式树 $TP(IC \cup r)$:

- 1) 如果 r 是一致的, 则 $TP(IC \cup r)$ 至少一个分支是开放的。
- 2) 如果开放 $TP(IC \cup r)$ 修复数据库 r , 则必然存在开放它的一个封闭分支修复数据库 r 。

证明: 1) 若 $TP(IC \cup r)$ 的每一个分支都是封闭的, 则 $TP(IC \cup r)$ 必然是一棵封闭树, 所以 r 是非一致的。根据逆否命题的等价性, 如果 r 是一致的, 则 $TP(IC \cup r)$ 必有一个分支是开放的。

2) 根据 1) 中 r 的修复实例 $r' \models IC$, 则 $TP(IC \cup r')$ 中必有一个分支是开放的, 所以要修复数据 r , 就必须开放 $TP(IC \cup r)$ 的分支。 r' 为对 $TP(IC \cup r)$ 开放任意一个分支得到的实例, 假设不存在 r' 是 r 的修复, 则 r' 是非一致数据库且 $TP(IC \cup r')$ 是开放的, 这与定义 7 相矛盾。因此开放 $TP(IC \cup r)$ 修复数据库 r , 则必然存在仅用开放它的一个封闭分支就能得到数据库实例 r 的修复。

定义 8 假设 $B=IC \cup r$ 是 $TP(IC \cup r)$ 的一个封闭分支, 其中 $I \in TP(IC)$:

- 1) 如果 I 不封闭, 也就是说分支 B 封闭仅取决于数据库文字, 则称 B 是数据封闭。
- 2) 令 $B=IC \cup r$ 是 $TP(IC \cup r)$ 的一个封闭分支, 定义 $op(B) := (r \setminus L(B)) \cup K(B)$, 其中:
 - (a) $L(B) = \{l \mid l \in r \text{ 且 } \neg l \in I\}$ 。
 - (b) $K(B) = \{l \mid l \text{ 是 } I \text{ 中的一个基原子并且不存在替换 } \sigma \text{ 使得 } l\sigma \in r\}$ 。

3) 对于 $TP(IC \cup r)$ 的一个数据封闭分支 B , 如果实例 r' 是实例 r 的一个开放当且仅当开放 $r' = op(B)$ 。

定理 2 非一致性数据库实例 r , 有 n 个不同的修复实例, 其中 r_i 表示 r 的第 i 个修复实例, 且 $r_i = (r \setminus L_i) \cup K_i$, 则:

- 1) $\Delta(r, r_i) = L_i \cup K_i$ 且 $L_i \cap K_i = \phi$ 。
- 2) $\bigcap_{i=1}^n L_i = \phi$, $\bigcap_{i=1}^n K_i = \phi$ 且 $\bigcap_{i=1}^n \Delta(r, r_i) = \phi (i > 1)$ 。
- 3) $\bigcap_{i=1}^n r_i \subseteq r (i > 1)$ 。

证明: 1) 因为 $\Delta(r, r_i) = (r - r_i) \cup (r_i - r)$ 且 $r_i = (r \setminus L_i) \cup K_i$, 所以 $L_i = r - r_i$ 且 $K_i = r_i - r$ 。因此得 $\Delta(r, r_i) = L_i \cup K_i$ 且 $L_i \cap K_i = \phi$ 。

2) (a) 假设当 $P(\bar{c}) \in \bigcap_{i=1}^n L_i$ 时, 根据定义 8, 则 $P(\bar{c}) \in r$ 且 $\neg P(\bar{c}) \in I$, 即在 $TP(IC \cup r)$ 的某一层的每个节点公式即为 $\neg P(\bar{c})$, 则完整性约束 $IC = \{\forall x \neg P(x)\}$ 或者 $IC = \{\exists x \neg P(x)\}$, 令 $I = \{P(x) \mid x \in D\}$ 。i) 对于前者 $IC = \{\forall x \neg P(x)\}$, 如果 $I \cap r \neq \phi$, 则只 r 有一个修复实例 $r' = r - \{P(x) \mid \sigma \in D \text{ 且存在替换 } \sigma \text{ 使得 } P(x)\sigma \in r\}$, 否则 $TP(IC \cup r)$ 是开放的, 即 r 是一致的; ii) 对于后者 $IC = \{\exists x \neg P(x)\}$, $TP(IC \cup r)$ 是开放的, 即 r 是一致的。以上两种情况均不成立, 故 $\bigcap_{i=1}^n L_i = \phi$ 。

(b) 假设当 $P(\bar{c}) \in \bigcap_{i=1}^n K_i$ 时, 根据定义 8, 则 $P(\bar{c}) \notin r$ 且 $P(\bar{c}) \in I$, 即在 $TP(IC \cup r)$ 的某一层的每个节点公式即为 $P(\bar{c})$, 则完整性约束 $IC = \{\forall x P(x)\}$ 或者 $IC = \{\exists x P(x)\}$, 令 $I = \{P(x) \mid x \in D\}$ 。i) 对于前者 $IC = \{\forall x P(x)\}$, 如果 $I \not\subseteq r$, 则 r 只有一个修复实例 $r' = r \cup \{P(x) \mid x \in D \text{ 且 } P(x) \notin r\}$, 否则 $TP(IC \cup r)$ 是开放的, 即 r 是一致的; ii) 对于后者 $IC = \{\exists x P(x)\}$, 如果 $I \cap r \neq \phi$, 则 $TP(IC \cup r)$ 是开放的, 即 r 是一致的, 否则修复实例 $r_i = r \cup \{P(i) \mid P(i) \in I\}$ 。以上两种情况均不成立, 故 $\bigcap_{i=1}^n K_i = \phi$ 。

(c) 因为 $\Delta(r, r_i) = L_i \cup K_i$ 且 $L_i \cap K_i = \phi$, 所以 $\bigcap_{i=1}^n \Delta(r, r_i) = (\bigcap_{i=1}^n L_i) \cup (\bigcap_{i=1}^n K_i)$ 。而 $\bigcap_{i=1}^n L_i = \phi$ 且 $\bigcap_{i=1}^n K_i = \phi$, 所以 $\bigcap_{i=1}^n \Delta(r, r_i) = \phi (i > 1)$ 。

3) 因为 $L_i \cap K_i = \phi$, 所以 $\bigcap_{i=1}^n r_i = (r / \bigcup_{i=1}^n L_i) \cup (\bigcap_{i=1}^n K_i)$ 。根

据 2), 当 $i > 1$ 时, 得 $\bigcap_{i=1}^n K_i = \phi$, 所以 $\bigcap_{i=1}^n r_i \subseteq r / \bigcup_{i=1}^n L_i$, 故 $\bigcap_{i=1}^n r_i \subseteq r (i > 1)$ 。

命题 1 非一致性数据库实例 $r(r| \neq IC)$ 和 $TP(IC \cup r)$ 上任一封闭分支 B , 如果 $r' = op(B)$, 则 $r'| = IC$ 。

证明: 设任意 $P(\bar{c}) \in \Delta(r, r')$ 。因为根据定义 8 有 $r' = op(B) = (r \setminus L) \cup K$ 且 $L \cap K = \phi$, 所以 $P(\bar{c}) \in L$ 或 $P(\bar{c}) \in K$ 。如果 $P(\bar{c}) \in L$, 即 $P(\bar{c}) \in r$ 且 $\neg P(\bar{c}) \in I$, 则 $P(\bar{c}) \notin r'$; 如果 $P(\bar{c}) \in K$, 即 $P(\bar{c}) \notin r$ 且 $P(\bar{c}) \in I$, 则 $P(\bar{c}) \in r'$ 。所以若 $\neg P(\bar{c}) \in I$, 则 $P(\bar{c}) \notin r'$; 若 $P(\bar{c}) \in I$, 则 $P(\bar{c}) \in r'$ 。故 r' 是开放的, 即 $r'| = IC$ 。

4.2 含有 I 封闭分支开放

定义 9 对于数据库实例 r 和公式树 $TP(IC \cup r)$, $TP(IC \cup r)$ 是数据封闭, 假设 ϕ 是 $TP(IC \cup r)$ 的任意一个结点:

1) 如果结点 ϕ 是 $TP(IC \cup r)$ 根结点, 则 ϕ 的结点封闭值 $V_r = 0$ 。

2) 如果结点 ϕ 是 $TP(IC \cup r)$ 非根结点, 则:

(a) 若结点 ϕ 是文字且封闭, 则 ϕ 的结点封闭值 V_r 等于 ϕ 父结点的结点封闭值加 1。

(b) 否则 ϕ 的结点封闭值 V_r 等于 ϕ 父结点的结点封闭值。

定义 10 对于数据库实例 r 和公式树 $TP(IC \cup r)$, $B = I \cup r$ 是 $TP(IC \cup r)$ 的任意一个封闭分支, 则分支 B 的封闭值 V_b 等于分支 B 的叶子结点封闭值 V_r , 数据库实例 r 的封闭值 $V = \min(V_r)$ 。

定理 3 给定一个数据库实例 r 和它的完整性约束 IC , 当且仅当对 $TP(IC \cup r)$ 中封闭值为 V 的分支 B 开放 $r' = op(B)$ 时, $r'| = IC$ 是数据库实例 r 的修复, 否则 r' 不是数据库实例 r 的修复。

证明: 对数据库实例 r 中任意一支 B 开放 $r' = op(B)$, 所以 $r'| = IC$ 且 $\Delta(r, r') = L \cup K$ 。 L 为从 r 中删除的文字集, K 为插入 r 中的文字集。根据定义 8, 删除或插入的文字都是使得分支 B 封闭的结点, 所以 L 和 K 中的文字集个数为 $|\Delta(r, r')| = |L| + |U| = V_b$ 。因为 $V_b \geq V$, 所以当且仅当对 $TP(IC \cup r)$ 中封闭值为 V 的分支 B 开放 $r' = op(B)$ 时, 得到的 $|\Delta(r, r')|$ 值最小, 则 $r'| = IC$ 是数据库实例 r 的修复。否则根据定义 4, r' 不是数据库实例 r 的修复。

推论 1 对于数据库实例 r 和它的公式树 $TP(IC \cup r)$, 经过结点 ϕ 的分支 B 的封闭值等于 ϕ 的结点封闭值加上分支 B 中属于 ϕ 的分支封闭值。

证明: 将分支 B 分为从根结点到结点 ϕ 和结点 ϕ 到该分支叶子结点两部分, 所以显然分支 B 的封闭值等于 ϕ 的结点封闭值加上分支 B 中属于 ϕ 的分支封闭值。

例 3 假设实例 $r, \Sigma(r) = \{P(a), R(b)\}$ 且 $IC = \{\forall x (P(x) \rightarrow Q(x))\}$, 它不满足 IC 。 $TP(IC \cup r)$ 如图 2 所示。

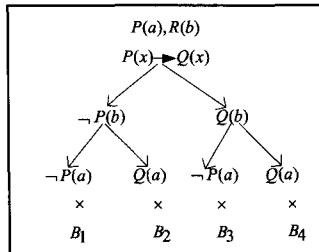


图 2 不满足 IC 的 $TP(IC \cup r)$

B_1, B_2, B_3 和 B_4 4 个分支的封闭值分别为 1、1、2 和 2, 而且通过分别开放 4 个分支 $r_i = op(B_i) (1 \leq i \leq 4)$ 得到的 4 个实例中仅有 r_1 和 r_2 是数据库实例 r 的修复, 我们可以通过开放分支封闭值最小的分支得到一个修复实例。

对于开放 $TP(IC \cup r)$ 的分支都是数据封闭的, 即 $TP(IC \cup r)$ 中的分支都不含有 I 封闭, 我们可以通过删除分支上导致数据库不一致的文字产生一个修复。因为完整性约束是严格的, 只允许改变数据库实例, 所以不能删除 I 中的文字。现在, 通过重新定义结点封闭值, 可以对含有 I 封闭的分支进行开放。

定义 11 (带有 I 封闭的结点封闭值) 对于数据库实例 r 的公式树 $TP(IC \cup r)$, 假设 ϕ 是 $TP(IC \cup r)$ 的任意一个结点, 如果 $\phi \in I$ 且 ϕ 封闭, 则结点的封闭值 V_r 等于 ϕ 父结点的结点封闭值加 2, 否则如定义 10。

推论 2 对于数据库实例 r 和含有 I 封闭的公式树 $TP(IC \cup r)$, 当封闭值为 V , 且任意叶子结点 ϕ 的结点封闭值 $V_r = V$ 时, 该叶子结点所在的分支 B 开放 $r' = op(B)$, 得到的 r' 是 r 的修复。

证明: 假设结点 $\phi \in I$ 是封闭的, 则与 ϕ 相邻的结点 ψ (ψ 不是 I 封闭) 的结点封闭值必然小于 ϕ 的结点封闭值; 又因为结点 ϕ 和 ψ 的子分支是相同的, 所以对于经过 ϕ 和 ψ 相同子分支的两个分支 B' 和 B'' , 根据推论 1 可得它们的分支封闭值 $V(B') = V(B'')$ 。所以对于任意叶子结点 ϕ , 如果它的结点封闭值 $V_r = V$, 则结点 ϕ 所在的分支 B 中的结点必然都不存在 I 封闭, 所以该叶子结点所在的分支 B 开放 $r' = op(B)$, 得到的 r' 是 r 的修复。

推论 3 对于非一致性数据库实例 $r(r| \neq IC)$ 和它的修复 r' , 数据库 r' 必然利用 $TP(IC \cup r)$ 封闭值开放分支得到。

证明: 当 r' 是 $r(r| \neq IC)$ 的一个修复实例时, 则必不存在另一个修复实例 r'' , 使得 $\Delta(r, r') \subseteq \Delta(r, r'')$, 即 $|\Delta(r, r')| < |\Delta(r, r'')|$ 。

假设 $r'' = op(B)$ 是 r 的一个修复实例且分支 B 的封闭值 $V_b > V$, 则 $\Delta(r, r'')$ 中文字个数 $|\Delta(r, r'')| = |L| + |U| = V_b$ 。现对 r 中任意一个分支封闭值为 V 的分支开放, 得到 $\Delta(r, r')$ 中文字个数 $|\Delta(r, r')| = |L| + |U| = V < V_b$ 。显然 r'' 不是 r 的修复, 所以对于数据库实例 r 的修复 r' , r' 必然是利用 $TP(IC \cup r)$ 封闭值开放分支得到。

5 一致性查询应答, CQA

定义 12 (一致性查询应答 CQA) 给定一个数据库实例 r, IC 集合和一阶查询 $Q(\bar{x})$, 如果一个基元组 $\bar{t}$ 是 Q 在 r 上的一致性查询应当且仅当对于 r 上的任意一个修复实例 r' 使得 $r'| = Q(\bar{t})$ 。

定理 4 对于数据库实例 r 和完整性约束 IC , 则对于其上的一阶查询 $Q(\bar{x})$ 能够获得一致性查询结果的范围集是 $\bigcap_{i=1}^n r_i$, 其中 r_i 表示 r 的第 i 个修复。

证明: 1) 当 $Q(\bar{x})$ 是一种带有自由变量 $\bar{x}$ 的查询时, 得到一致性查询结果为 $Q(\bar{t})$, 则对于数据库实例 r 的任一修复 r_i , 必须使得 $Q(\bar{t}) \in r_i (1 \leq i \leq n)$, 所以 $Q(\bar{t}) \in \bigcap_{i=1}^n r_i$ 。

2) 当 Q 是不带自由变量 $\bar{x}$ 的真值查询, 且一致性查询结果 Q 为真时, 对于数据库实例 r 的任一修复 r_i , 必须使得 $Q \in r_i (1 \leq i \leq n)$, 所以 $Q \in \bigcap_{i=1}^n r_i$ 。

(下转第 208 页)

- tion methods based on linear programming[J]. Pattern Recognition, 2011, 44(1): 97-106
- [2] Kittler J, Hatef M, Duin R P W, et al. On combining classifiers[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(3): 226-239
 - [3] Zhou Z-H, Yu Y. Ensembling local learners through multimodal perturbation [J]. IEEE Trans. SMC—Part B: Cybernetics, 2005, 35: 725-735
 - [4] Breiman L. Bagging Predictors[J]. Machine Learning, 1996, 24(2): 123-140
 - [5] Schapire R E. The strength of weak Learnability[J]. Machine Learning, 1990, 5(2): 197-227
 - [6] Hu Q-H, Yu D-R, Xie Z-X, et al. EROS: ensemble rough subspaces[J]. Pattern Recognition, 2007, 40: 3728-3739
 - [7] Hu Q-H, Yu D-R, Xie Z-X. Neighborhood classifiers[J]. Expert Systems with Applications, 2008, 34: 866-876
 - [8] Valentini G, Masulli F. Ensembles of learning machines[C]// Proc of Valentini02. Neural Nets WIRN Vietri-02, Series Lecture Notes in Computer Sciences. Springer-Verlag, 2002: 3-19
 - [9] Rosset S, Zhu J, Hastie T. Boosting as a Regularized Path to a Maximum Margin Classifier[J]. Journal of Machine Learning Research, 2004, 5: 941-973
 - [10] 万小毛, 鲍明, 赵淳生. 齿轮箱故障诊断技术综述[J]. 振动、测试与诊断, 1990, 10(4)
 - [11] Kim S-J, Koh K, Lustig M, et al. A method for large-scale l1-regularized least squares[J]. IEEE Journal on Selected Topics in Signal Processing, 2007, 1(4): 606-617
 - [12] Lei Y G, Zuo M J. Gear crack level identification based on weighted k neighbor classification algorithm [J]. Mechanical Systems and Signal Processing, 2009, 23(5): 1535-1547
 - [13] Fumera G, Roli F. A theoretical and experimental analysis of linear combiners for multiple classifier systems[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(6): 942-956
 - [14] Garg A, Roth D. Margin distribution and learning[C]// Proc. Int. Conf. Mach. Learn., Washington, DC, 2003: 210-217
 - [15] Schapire R E, Bartlett P, Freund Y, et al. Boosting the margin: A new explanation for the effectiveness of voting methods[J]. Annals of Statistics, 1998, 26(5): 1651-1686

(上接第 180 页)

因此, 能够获取一致性查询结果的数据库范围集为 $\bigcap_{i=1}^n r_i$ 。

例 4 (例 2 续) 考虑完整性约束: $IC = \{ \forall (x, y, z) (Supply(x, y, z) \wedge Product(z, T_4) \rightarrow x = C) \}$ 和非一致数据库实例 $r = \{ Supply(C, D_1, It_1), Supply(D, D_2, It_2), Product(It_1, T_4), Product(It_2, T_4) \}$ 。

数据库实例 r 仅有两个修复实例分别是 $r' = \{ Supply(C, D_1, It_1), Product(It_1, T_4), Product(It_2, T_4) \}$ 和 $r'' = \{ Supply(C, D_1, It_1), Supply(D, D_2, It_2), Product(It_1, T_4) \}$ 。第一种查询“ $Supply(x, y, z)?$ ”, 元组 (C, D_1, It_1) 是一致性结果, 因为它能够从两个修复实例中被获取, 但是 (C, D_2, It_2) 不能从 r' 中获取, 所以它不是一致性结果, 则查询结果 $Q(i) \in (\bigcap_{i=1}^n r_i = \{ Supply(C, D_1, It_1), Product(It_1, T_4) \})$ 。第二种查询“ $Supply(C, D_1, It_1)?$ ”结果为真, 而“ $Supply(D, D_2, It_2)?$ ”为假; 若查询为真, 则 $Q(i) \in r_i (1 \leq i \leq n)$, 即 $Q(i) \in (\bigcap_{i=1}^n r_i = \{ Supply(C, D_1, It_1), Product(It_1, T_4) \})$ 。一致性查询结果必须都能从它的每一个修复实例中获取到。

结束语 数据库完整性约束 IC 作为数据库模式的一部分, 有效地保证了数据的完整性和有效性, 使数据符合现实世界的实体规则。现有的商业 DBMS 为 IC 提供了支持, 但总聚焦在发展一系列的约束关系来尽可能保证每一个数据库是合法和一致的。然而, 现实世界的一个实体在数据库中常常是对应多个不一致的数据。本文利用 $TP(IC \cup r)$ 分支封闭和开放规则, 将非一致性数据库与 tableau 推理相结合, 根据结点封闭值定义来计算确定分支封闭值, 然后通过开放极小分支封闭值的分支来实现非一致性数据库的修复, 并将这种方法扩展到带有 I 封闭的非一致性数据库。这样可以避免通过清除方法实现修复从而导致数据信息丢失的问题。这种方法对于实现数据库修复和保证数据库的完整性具有极其重要的意义。未来, 我们考虑将该方法与数据库一致性查询相结合, 来提高非一致性数据库的一致数据重写查询性能。

参考文献

- [1] Bertossi L. Database repairing and consistent query answering [M]. Morgan & Claypool publishers, 2011
- [2] Arenas M, Bertossi L, Chomicki J. Consistent query answers in inconsistent databases[C]// ACM Symposium on Principles of Database Systems (ACM PODS'99). ACM Press, 1999: 68-79
- [3] Bertossi L, Schwind C. An analytic tableaux based characterization of database repairs for consistent query answering (preliminary report)[C]// Working Notes of the IJCAI'01 Workshop on Inconsistency in Data and Knowledge. AAAI Press, 2001: 96-106
- [4] 李娇, 刘全, 傅启明, 等. 分布式数据库中基于局部 CON 模型的记录匹配方法[J]. 通信学报, 2011, 32(7): 196-202
- [5] Fan W, Geerts F, Ma S, et al. Detecting inconsistencies in distributed data[C]// the IEEE International Conference on Data Engineering (ICDE). 2010: 64-75
- [6] Fan W, Jia X, Li J, et al. Reasoning about record matching rules [J]. the International Conference on Very large Data Bases (VLDB), VLDB Endowment, 2009, 2(1): 407-418
- [7] Bertossi L. Consistent query answering in databases[J]. ACM SIGMOD Record, 2006, 35(2): 68-79
- [8] Chomicki J, Marcinkowski J, Staworko S. Computing consistent query answers using conflict hypergraphs[C]// ACM International Conference on Information and Knowledge Management. Washington DC: ACM Press, 2004: 417-426
- [9] Elmagarmid A K, Ipeirotis P G, Verykois V S. Duplicate record detection: a survey[J]. IEEE Transaction on Data and Knowledge Engineering (TKDE), 2007, 19(1): 1-16
- [10] 刘全, 伏玉琛, 孙吉贵, 等. 一种基于符号集合的自动推理扩展方法[J]. 计算机研究与发展, 2007, 44(8): 1317-1323
- [11] 刘全, 孙吉贵, 崔志明. 基于布尔剪枝的多值广义量词 Tableau 推理规则简化方法[J]. 计算机学报, 2005, 28(9): 1514-1518
- [12] Melvin F. First-order logic and automated theorem proving[M]. New York: Springer Verlag, 1996
- [13] Golab L, Karloff H, Korn F, et al. On generating near optional tableaux for conditional functional dependencies[J]. the International Conference on Very large Data Bases (VLDB), VLDB Endowment, 2008, 1(1): 376-390