一种基于语义 tableau 的数据库修正方法*)

刘 全[1,2] 凌兴宏[1] 张宏斌[1] 孙吉贵[2]
(苏州大学计算机科学与技术学院 苏州 215006)
(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 tableau 作为自动推理的有效方法之一在许多人工智能领域中有重要的应用。在 tableau 基础上,提出新的 tableau 开放和封闭的推理标准,应用于数据库实例不满足完整性约束的不相容关系数据库中,并对其进行修正。这样可以采用逻辑程序的方法,对数据库进行修正,解决了传统修正方法丢失信息、出现新的不相容等问题。

关键词 tableau,完整性约束,数据库修正

A Method of Database Repairs Based on Semantic Tableau

LIU Quan[1,2] LING Xing-Hong[1] ZHANG Hong-Bin[1] SUN Ji-Gui[2]
(Institute of Computer Science and Technology, Soochow University, Suzhou 215006)[1]
(Laboratory of Symbol Calculation and Knowledge Engineering, Ministry of Education, Jilin University, Changchun 130012)[2]

Abstract As one of effective automated reasoning methods, tableau has been applied to many important AI fields. On the base of tableau, an open and close new reasoning criterion is proposed. The criterion is applied to inconsistent relational database which database instances can not satisfy integrity constrains. Database can be repaired through use logic programming, which can solve some problems such as lose information, arise new inconsistent.

Keywords Tableau, Integrity constraint, Database repairs

1 引言

自动推理作为自动定理证明的扩展是人工智能研究的基础工作,许多重要的人工智能系统都是以推理为其核心部分。其中的 tableau 方法,由于具有通用性、直观性及易于计算机实现等特点,已经成为重要的自动推理方法之一,并在形式校验、自然语言处理、模糊控制、专家系统、数据库及知识表示等许多领域得到较好的应用。

完整性约束是保证数据库一致性的重要概念。任何一个数据库都会由于某些自然因素或人为因素而受到局部或全局的破坏,经常出现完整性得不到满足,产生不相容数据库的情况[1]。如在数据仓库中,由于数据来源不同,因此无法保证所有的数据库实例都满足完整性约束。另外在关联数据库中,完整性约束在一个数据库中满足但在另一个关联数据库中却无法满足。这些因素可能是系统设备故障或软件的错误,也可能是应用程序编制上的错误或操作员的操作错误而引起的,这些因素有的是不可避免的。解决这些问题的传统方法是从数据库中删除不相容信息,这样一方面浪费大量的人力物力,更重要的是,由于判断不当,会导致有用信息丢失,有时甚至会出现新的不相容现象。因此,如何及时发现并采取措施防止错误进一步蔓延并能及时恢复,对于一个管理信息系统来说是非常重要的。本文在传统 tableau 基础上,提出新的 tableau 开放和封闭的推理标准,应用到数据库实例不满足完整性约束的不相容关系数据库中,并对其进行修正[2]。这样可以采用逻辑程序的方法,对数据库进行修正,解决了传统修正方法存在的一些问题。为了节省篇幅,这里使用的未解释的记号和概念以及 tableau 推理的基本知识,请参见文[3～5]。

2 语义 tableau

一个公式集的 tableau 过程是通过递归地使用 tableau 规则将公式拆分为子公式而得到的公式集的集合,tableau 推理规则如表 1 所示。α-规则向分支中加入新公式,β-规则将公式分成两个部分并加入到 tableau 分枝中。给定一个公式 ϕ,用 TP(ϕ)表示通过 tableau 系统产生的公式树,公式树可以被看成是表示为 $X,Y,\cdots$ 的分枝的集合。

表 1 Tableau 推理规则

α	α1	α2	β	β1	β2	δ	δ(p)	γ	γ(p)
$f_1\wedge f_2$	f_1	f_2	$f_1\vee f_2$	f_1	f_2	$(\exists x)f$	$f\{p/x\}$	$(\forall x)f$	$f\{p/x\}$
$\neg(f_1\vee f_2)$	$\neg f_1$	$\neg f_2$	$\neg(f_1\wedge f_2)$	$\neg f_1$	$\neg f_2$	$\neg(\forall x)f$	$\neg f\{p/x\}$	$\neg(\exists x)f$	$\neg f\{p/x\}$
$\neg(f_1\rightarrow f_2)$	f_1	$\neg f_2$	$f_1\rightarrow f_2$	$\neg f_1$	f_2	P是新常量		P是任意常量	

一个 tableau 分枝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。每个开放分枝对应着一个公式的模

*)本课题得到国家自然科学基金(60273080,60473003)资助。刘 全 博士,副教授,研究方向为智能信息处理,自动推理,地理信息系统。孙吉贵 博士,教授,博士生导师,研究方向为人工智能,自动推理。

型：如果一个分枝 B∈TP(ϕ)是开放的，那么在 B 上的基原子集是 ϕ 的一个模型。如果最初的公式集是不相容的，那么它没有模型，并且所有的分枝都是封闭的。存在完备性的定理：F 是定理当且仅当 TP(¬F)是封闭的。TP 不只是公式的定理证明器，而是一个公式集转化为树的一个应用，因此在 tableau 上的操作，能够看成以下形式：

定义 1 令 ϕ 和 ψ 是任意公式，那么 TP 有下列性质：

(1)$TP(\{\phi\vee\psi\})=TP(\{\phi\})\cup TP(\{\psi\})$

(2)$TP(\{\phi\wedge\psi\})=\{X\cup Y: X\in TP(\{\phi\})$ 且 $Y\in TP(\{\psi\})\}$

(3)如果 $B\in TP(\{\phi\wedge\psi\})$，那么 $B=B^{*}\cup B^{**}$，$B^{*}\in TP(\{\phi\})$ 且 $B^{**}\in TP(\{\psi\})$。

在定义 1 中，性质(3)来自于性质(1)和(2)，由定义 1 可以得出如下定义：

定义 2 给定 tableau T 和 T^{*}，X,Y 是 T 和 T^{*} 的分枝集，那么 tableau 的结合 $T\oplus T^{*}=\{X\cup Y: X\in T$ 且 $Y\in T^{*}\}$。

由定义 2 可知，检查 $TP(\{\neg A\rightarrow\phi\})$是否封闭时，只要检查 $TP(\{A\})T\oplus TP(\{\neg\phi\})$是否是封闭的。

3 语义 tableau 表示数据库实例

为了使用语义 tableau 表示数据库实例和查询相容答案，提出了一种特殊形式的 R-tableau，这种 R-tableau 适合于表示数据库实例和它们的完整性约束。

给定一个数据库实例 r 和一个完整性约束集合 IC。对于 IC 和 r 的 tableau，$TP(IC\cup r)$是一个以公式集 $IC\cup r$ 为根节点的推理树，每个 tableau 分枝 B 为 $I\cup r$ 形式，$I\in TP(IC)$，I 称为分枝的 IC 部分，如果 r 是不相容的，tableau 是封闭的，即 tableau 存在封闭分枝。由于 IC 是相容的，因此一个 tableau 的 IC 部分永远是不封闭的，只有 r 和 IC 相结合才能产生一个封闭的 tableau。通过从封闭的分枝中删除数据库文字可以将 r 转换成数据库实例并且得到一个数据库修正。

在数据库理论中，通常作如下假设：(1)唯一名称假设(Unique Names Assumption, UNA)：如果 a 和 b 在 D 中是不同的常量，那么在 r 中 $a\neq b$ 成立。(2)封闭世界假设(Closed World Assumption, CWA)：如果 r 是一个数据库实例，那么对于任何数据库基原子 $P(c)$，如果 $P(c)\notin r$，$\neg P(c)\in r$。

当计算一个关于数据库实例 r 的 tableau 时，需要将 UNA 和 CWA 的假设加入其中，这样需要在 tableau 方法基础上，增加 tableau 封闭条件。例如对于 D 中的不同常量 a 和 b，当 tableau 分枝中出现公式 $a=b$ 时，这个分枝即封闭。

例 1 考虑完整性约束

IC: $\forall x,y,z(Material(x,y,z)\wedge Product(y,PAPER)\rightarrow x=\text{No0001})$

表明在数据库中原料 WOOD 是产品 PAPER 的唯一提供者。假设有一个不相容的数据库实例 r：

Material			Product	
No0001	WOOD	WHITE	WOOD	PAPER
No0002	STEEL	BLACK	STEEL	PAPER

在这种情况下，$TP(IC\cup r)$为如图 1 所示的推理树。

```
∀x, y, z(Material(x, y, z)∧Product(y, PAPER)→x=No0001)
Material(No0001, WOOD, WHITE)
Material(No0002, STEEL, BLACK)
Product(WOOD, PAPER)
Product(STEEL, PAPER)
Material(No0001, WOOD, WHITE)∧Product(WOOD, PAPER)→No0001=No0001
Material(No0002, STEEL, BLACK)∧Product(STEEL, PAPER)→No0002=No0001
¬Material(No0001, WOOD, WHITE) | ¬Product(WOOD, PAPER) | No0001=No0001
              *                |            *          |
     ¬Material(No0002, STEEL, BLACK) | ¬Product(STEEL, PAPER) | No0002=No0001
                   *                 |           *            |       *
```

图 1 $TP(IC\cup r)$的推理树

从图 1 中可以看出，$TP(IC\cup r)$是封闭的，r 对于 IC 是不相容的。节点 Material(No0001, WOOD, WHITE) ∧ Product(WOOD, PAPER)→No0001=No0001 和 Material(No0002, STEEL, BLACK) ∧ Product(STEEL, PAPER)→No0002=No0001 是对公式 $\forall x,y,z(Material(x,y,z)\wedge Product(y,PAPER)\rightarrow x=\text{No0001})$ 应用 γ-规则得到的，对 Material(No0002, STEEL, BLACK) ∧ Product(STEEL, PAPER)→No0002=No0001 应用 β-规则产生三个子树：¬Material(No0002, STEEL, BLACK)，¬Product(STEEL, PAPER)，No0002=No0001。整个 tableau 封闭。

在通常情况下，当在一个 tableau 分枝中包含一个公式和它的否定时即封闭。然而在此需要考虑 UNA 和 CWA 两种情况。因此对标准的 tableau 封闭条件修改如下。

定义 3 令 B 是一个带完整性约束 IC 的数据库实例 r 的 tableau 分枝，即 $IC\cup r$。B 是封闭的当且仅当下列条件之一满足：

(1)在 D 中对于不同的常量 a 和 b，$a=b\in B$。

(2)①对于在 D 中的元素组成的基原子 c，$P(c)\in I$ 且 $P(c)\notin r$。

②$P(c)\in I$ 并且在 c 中不存在替换 σ 使得 $P(c)\sigma\in r$。

(3)对于在 D 中的元素组成的基原子 c，$\neg P(c)\in I$ 且 $P(c)\in r$。

(4)对于任意公式 ϕ，$\phi\in B$ 且 $\neg\phi(\in B$。

(5)对于任意项 t，$\neg t=t\in B$。

条件(1)考虑 UNA 假设，条件(2)考虑的是 CWA 假设。

例 2 考虑完整性约束 IC: $\forall x(P(x)\rightarrow\exists yQ(x,y))$和一个数据库实例 $r=\{P(a),Q(b,d)\}$，$a、b、c\in D$，以 $TP(IC\cup r)$为根的 tableau 如图 2 所示。

```
∀x(P(x)→∃yQ(x, y))
P(a), Q(b, d)
P(a)→Q(a, f(a))
¬P(a) | Q(a, f(a))
  *   |     *
```

图 2 不相容的 tableau

由于 $Q(a,f(a))$ 不属于数据库实例，因此右侧分枝封闭，在活动的数据库域中不存在 x 使得 r 中包含 $Q(a,x)$，根据 CWA 假设，对于任意 x，r 中包含 $\neg Q(a,x)$，因此分枝中包含 $Q(a,f(a))$ 封闭，r 是关于 IC 不相容的。

将数据库实例改成 $\{P(a),Q(a,d)\}$，它对于同一 IC 是相容的。可以得到如图 3 所示的 tableau 过程。

$$\forall x(P(x)\rightarrow \exists y Q(x,y))$$
$$P(a),Q(a,d)$$
$$P(a)\rightarrow Q(a,f(a))$$
$$\neg P(a) \mid Q(a,f(a))$$
$$* \qquad *$$

图 3 修正后的相容 tableau

可以定义 $f(a)=d$，使得 $Q(a,f(a))$ 成为数据库中的一个成员，修正后的 tableau 右侧分枝是开放的。

引理 1 令 S 为公式集，并且 γ 和 δ 为公式。

1) 如果 $S\cup\{\gamma\}$ 是可满足的，那么 $S\cup\{\gamma,\gamma(t)\}$ 对于任意封闭项 t 也是可满足的。

2) 如果 $S\cup\{\delta\}$ 是可满足的，那么 $S\cup\{\delta,\delta(p)\}$ 对于任意对 S 和 δ 是新的常量符号 p 也是可满足的。

证明：1)假设在模型 $M=\langle D,I\rangle$ 中，$S\cup\{\gamma\}$ 是可满足的。只要证明在同一模型下 $S\cup\{\gamma,\gamma(t)\}$ 是可以满足的。

因为 γ 在模型中为真，所以 $(\forall x)\gamma(x)$ 也为真，这里 x 对 γ 来说也是新变量，那么对任何赋值 A，$[\gamma((x)]^{I,A}$ 为真。现在，令 A 为一个赋值，使得 $x^A=t^\gamma$，可得 $[\gamma(t)]^{I,A}=[\gamma\{x/t\}]^{I,A}=[\gamma(x)]^{I,A}=t$，即 $S\cup\{\gamma,\gamma(t)\}$ 在模型 $M=\langle D,I\rangle$ 中是可以满足的。

2)假设在模型 $M=\langle D,I\rangle$ 中，$S\cup\{\delta\}$ 是可满足的，p 是对 S 和 δ 为新的常量符号，只要证明在任意模型下 $S\cup\{\delta,\delta(p)\}$ 是可以满足的。

因为 δ 在模型中为真，所以 $(\exists x)\delta(x)$ (x 对于 δ 是新的)且 $[\delta(x)]^{I,A}$ 对于某一赋值 A 也为真。

构造新的模型 $M^*=\langle D,J\rangle$，有同样的定义域，一个解释 J 除了 p 外与 I 完全相同，作为特例设 $p^J=x^A$，因此这两个模型只是在常量符号 p 处不同，这样 $S\cup\{\delta\}$ 在模型 M^* 中是可以满足的，且 $[\delta(x)]^{J,A}$ 也是真的。又因 $x^A=p^J$，有 $[\delta(x)]^{J,A}=[\delta\{x/p\}]^{J,A}=[\delta(x)]^{J,A}=t$。因此 $S\cup\{\delta,\delta(p)\}$ 在 M^* 中是可以满足的，因此在 M 中也是可以满足的。

引理 2 如果任何 tableau $TP(IC\cup r)$ 扩展规则被应用到可满足的表，其结果为另一个可满足的表。

证明： 假设 T 是一个可满足的 tableau $TP(IC\cup r)$，对于 T 的分枝 θ 应用表扩展规则产生 X，生成一个 tableau T^*。下面证明 T^* 也是可以满足的。

因为 T 是可以满足的，T 至少有一个可满足的分枝，选择这一分枝，记为 τ。

1) $\tau\neq\theta$。规则应用到 θ 上，τ 仍为 T^* 的分枝，因此 T^* 是可以满足的。

2) $\tau=\theta$。θ 本身是可满足的，布尔赋值 ν 将 θ 中的所有公式映射为 t。

2a) $X=\alpha$。那么 θ 被扩展为 α_1 和 α_2 而产生 T^*。由于 α 产生在 θ 上，有 $\nu(\alpha)=t$，$\nu(\alpha)=\nu(\alpha_1)\wedge\nu(\alpha_2)$，因此 ν 必将 $\nu(\alpha_1)$ 和 $\nu(\alpha_2)$ 都映射为 t。这样 ν 将在 T^* 的 θ 扩展的每个公式都映射为 t，因此 T^* 是可以满足的。

2b) $X=\beta$。那么将左右儿子加到 θ 的最后节点上，标记为 β_1 和 β_2，而产生 T^*，由于 β 产生在 θ 上，$\nu(\beta)=t$，但 $\nu(\beta)=\nu(\beta_1)\vee\nu(\beta_2)$，$\nu$ 将 β_1 或 β_2 映射为 t，ν 将左侧或右侧的扩展分枝映射为 t。只要存在其一，T^* 就有可满足的分枝，因此 T^* 是可以满足的。

其它情况：X 是 $\neg\neg Z$，$\neg\top$ 或 $\neg\bot$，易证 T^* 是可以满足的。

另外，含量词的二种情况，引理 1 已证。

定理 1 对于一个数据库实例 r 和完整性约束 IC，如果 $TP(IC\cup r)$ 是封闭 tableau 的，那么 r 是关于 IC 不相容的，即 r 不满足 IC。

证明：假设公式集 $S=IC\cup r$ 存在一个 tableau，S 是可满足的。

对 S 构造一个开始于 S 封闭的 tableau，由于 S 是可满足的，那么初始标记为 S 的 tableau 是可满足的，根据引理 2，S 的 tableau 的每个子 tableau 都是可满足的，包括最后封闭的 tableau，这显然是矛盾的，因此，S 是不可满足的，即 $IC\cup r$ 是不可满足的，那么 r 是关于 IC 不相容的，即 r 不满足 IC。

结论 本文给出了一种对不相容数据库修正的逻辑方法，利用该方法对数据库修正，可以避免传统方法信息丢失等问题。这种逻辑修正方法可以应用到数据仓库和数据挖掘中，对于保证数据库完整性具有重要的意义。

参 考 文 献

1 Arenas M, Bertossi L, Chomicki J. Consistent query answers in inconsistent databases. In: Proceeding ACM Symposium on Principles of Database Systems(ACM PODS'99), ACM Press, 1999. 68～79

2 Greco S, Zumpano E. Computing repairs for inconsistent databases. In: Proceedings 3th Intl. Symposium on Cooperative Database Systems for Advanced Applications(CODAS01), Beijing, China, 2001

3 刘全，孙吉贵. 提高一阶多值逻辑 tableau 推理效率的布尔剪枝方法. 计算机学报，2003，26(9)：1165～1170

4 Fitting M C. First-order logic and automated theorem proving. New York: Springer Verlag, 1996

5 Fitting M C. Types and Tableau. New York: Springer Verlag, 2000

6 Davey B A, Priestley H A. Introduction to lattices and order. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1990

7 刘全，孙吉贵. 非经典逻辑的语义 tableau 方法. 计算机科学，2002，29(5)：72～75

8 Bertossi L, Schwind C. Analytic tableaux and database repairs. In: Foundations of Information and Knowledge Systems, Springer LNCS 2284, 2002

9 刘全，孙吉贵，于万钧. 自由变量语义 tableau 中(-规则的一种改进方法. 计算机研究与发展，2004，41(7)：1068～1073

(上接第 75 页)

4 胡宝清. 模糊理论基础[M]. 武汉：武汉大学出版社，2004

5 许树柏. 层次分析法原理[M]. 天津：天津大学出版社，1987

6 魏毅强，刘进生，王绪柱. 不确定型 AHP 中判断矩阵的一致性概念及权重. 系统工程理论与实践，1994，14(4)：16～22

7 尤天慧，樊治平. 不确定性多属性决策中确定熵权的一种误差分析方法. 系统工程，2003，21(1)：101～104

8 朱孔来. 评价指标的非线性无量纲模糊处理方法. 系统工程，1996，14(6)：58～62

9 Ellison R J, Fisher D A, Linger R C, et al. An Approach to Survivable Systems. http://www.cert.org/easel/natol.doc