

一种基于语义 Tableau 的不相容知识处理方法

刘全^{1,2}, 伏玉琛¹, 凌兴宏¹, 孙吉贵²

(1. 苏州大学计算机科学与技术学院, 苏州 215006; 2. 吉林大学符号计算与知识工程教育部重点实验室, 长春 130012)

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, 该文提出了 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。采用了逻辑程序的方法, 对数据库进行修正, 解决了传统修正方法丢失信息、出现不相容等问题。

关键词: Tableau; 完整性约束; 不相容知识处理

Method of Inconsistent Knowledge Processing Based on Semantic Tableau

LIU Quan^{1,2}, FU Yuchen¹, LING Xinghong¹, SUN Jigui²

(1. Institute of Computer Science and Technology, Soochow University, Suzhou 215006;

2. Key Laboratory of Symbol Calculation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012)

【Abstract】 As one of effective automated reasoning methods, Tableau is 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 logic programming, which can solve some problems such as losing information, arising new inconsistent.

【Key words】 Tableau; Integrity constraint; Inconsistent knowledge processing

1 概述

由于现实世界中的信息往往是不确定的, 人们认识事物的过程也是不断地发展变化的, 各人有不同的主观观点, 对事物的观察也有一定的偏差, 甚至有些情况下, 还不得不采取估计和猜测的方法, 因此在知识表达系统中往往存在一定程度的不相容性。这种不相容性产生的主要原因在于^[1]: 信息表中可能存在遗失数据, 对其进行补齐(完备化), 需要通过现实事物的猜测信息; 在科学实验中, 由于一些观察失误或者观察不足, 可能获取到一些具有冲突的信息; 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表中的实例都满足完整性约束; 在关联知识库中, 完整性约束在一个知识库中满足但在另一个关联知识库中却无法满足。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, 有些是不可避免的。解决这些问题的传统方法是手工处理知识库中不相容信息, 这样不仅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。如何及时发现并采取措施, 防止错误进一步蔓延并能得到及时恢复, 对于一个基于知识的系统来说是非常重要的。本文在传统 Tableau 基础上, 提出了新的 Tableau 开放和封闭的推理标准和策略, 应用到知识库实例不满足完整性约束的不相容关系知识库中, 并对其进行处理^[2]。这样可以采用逻辑程序的方法, 对知识库中的知识进行处理。使用该方法处理不相容知识, 一方面解决了传统处理方法存在的一些问题, 另一方面将一些经典逻辑中识别 γ 公式^[3]、改进 δ 公式^[4], 减少证明空间和时间、提高推理效率的方法加入其中, 可以提高处理海量数据的速度。为了节省篇幅, 这里使用的未解释的

记号、概念、Tableau 推理的基本知识, 请参见文献[5~7]。

2 知识库实例的表示

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, $TP(IC, r)$ 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, $I \subseteq TP(IC, r)$, 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

基金项目: 国家自然科学基金资助项目(60273080, 60473003)

作者简介: 刘全(1969-), 男, 博士、副教授, 主研方向: 智能信息处理, 自动推理, 地理信息系统; 伏玉琛, 博士、副教授; 凌兴宏, 博士、讲师; 孙吉贵, 博士、教授、博导

收稿日期: 2006-05-08 **E-mail:** quanliu@suda.edu.cn

Tableau 分枝中出现公式 $a=b$ 时, 这个分枝即封闭。

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。由于在此需要考虑 UNA 和 CWA 2 种情况, 因此对标准的 Tableau 封闭条件修改如下。

定义 令 B 是一个带完整性约束 IC 的数据库实例 r 的 Tableau 分枝, 即 $IC \vdash r$ 。 B 是封闭的, 当且仅当满足下列条件之一:

- (1) 在 D 中对于不同的常量 a 和 b , $a=b \notin B$ 。
- (2) 对于在 D 中的元素组成的基原子 $c, P(c) \in B$ 且 $P(c) \notin r$;
- (3) 对于在 D 中的元素组成的基原子 $c, \neg P(c) \in B$ 且 $P(c) \in r$ 。
- (4) 对于任意公式 $\phi, \phi \in B$ 且 $\neg \phi \in B$ 。
- (5) 对于任意项 $t, \neg t \in B$ 。

条件(1)考虑 UNA 假设, 条件 (2) 考虑的是 CWA 假设。

例 考虑完整性约束 IC 为

$$\forall x(P(x) \rightarrow \exists yQ(x,y))$$

一个数据库实例为

$$r = \{P(a), Q(b,d)\}, a, b, c \in D$$

以 $TP(IC \vdash r)$ 为根的 Tableau 如图 1 所示。

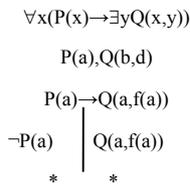


图 1 不相容的 Tableau

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 2 所示 Tableau 过程。

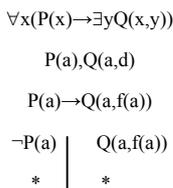


图 2 修正后的相容 Tableau

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 $x^A = t^i$, 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, 由于一个解释 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。又由于 $x^A = p^J$, 有 $[\delta(x)]^{J,A} = [\delta\{x/p\}]^{J,A} = [\delta(x)]^{I,A} = t$, $S \cup \{\delta, \delta(p)\}$ 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 $TP(IC \vdash r)$ 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 $TP(IC \vdash r)$, 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, 所以 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, T^* 是可以满足的。

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

1) $X = \alpha$ 。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 $v(\alpha) = t$, $v(\alpha) = v(\alpha_1) \vee v(\alpha_2)$, 因此 v 必将 $v(\alpha_1)$ 和 $v(\alpha_2)$ 都映射为 t 。这样 v 将在 T^* 的 θ 扩展的每个公式都映射为 t , T^* 是可以满足的。

2) $X = \beta$ 。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, $v(\beta) = t$, 但 $v(\beta) = v(\beta_1) \vee v(\beta_2)$, v 将 β_1 或 β_2 映射为 t , v 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。

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

另外, 对于含量词的 2 种情况, 引理 1 已证明。

定理 对于一个数据库实例 r 和完整性约束 IC, 如果 $TP(IC \vdash r)$ 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 $S = IC \vdash r$ 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, 由于 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 $IC \vdash r$ 是不可满足的, 那么 r 是关于 IC 不相容的, 即 r 不满足 IC。

3 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[C]//Proc. of ACM Symposium on Principles of Database Systems. 1999: 68-79.
- 2 Greco S, Zumpano E. Computing Repairs for Inconsistent Databases[C]//Proc. of the 3th International Symposium on Cooperative Database Systems for Advanced Applications. 2001.

(下转第 177 页)