

一种基于三值逻辑的关系演算^{*}

A Relational Calculus Based on Three-Valued Logic

毛宇光^{1,2} 徐洁磐¹ 韩 波^{2,3}

(南京大学计算机软件新技术国家重点实验室 南京210093)¹

(南京航空航天大学信息科学与技术学院 南京210016)²

(南京航空航天大学计算机科学研究所 南京210016)³

Abstract Relational calculus is one of theoretical foundations for the relational databases. The classical relational calculus based on two-valued logic only applies to relational databases with complete information. Under null circumstance, relational database systems usually adopt three-valued logic to deal with missing information. To meet the needs of handling incomplete information, this paper presents a three-valued tuple relational calculus MFMTRC and a three-valued domain relational calculus MFMDRC.

Keywords Incomplete information, Three-valued logic, Null, Relational calculus

1 引言

不完全信息数据库是数据库理论的一个重要研究方向^[1~5]。现实世界中不完全信息是普遍存在的。数据库是对客观世界的模拟,必须考虑不完全信息的处理。关系数据库中一般采用空值来表示不完全信息,由于空值的存在,比较运算和逻辑运算的结果可能为真(T)、假(F)和未知(U)三个值,经典的以二值逻辑为基础的关系数据库理论难以适应这种情况^[1,3]。为适应不完全信息的处理,基于二值逻辑的关系演算必须扩充。本文在三值谓词演算系统 MFM^[6]的基础上,提出了一种表达能力较强的三值元组关系演算 MFMTRC 和三值域关系演算 MFMDRC。

2 缺失信息

缺失信息(Missing Information)的引入与处理已成为关系数据库理论的一个重要课题,关系模型应提供一种合理机制处理这个问题。信息缺失的主要原因有值未知、值不存在、值不可应用、值未定义等。有两种处理缺失信息的方法:

①Codd 的方法 插入某种类型的特殊标志--NULL(空值)来指明属性的值是缺失的;

②Date 的方法 插入某种类型的缺省值(default)。

由于目前主流 DBMS 的实现均采用第一种方法来处理缺失信息,因此着重讨论 Codd 的方法。通常来讲,空值可按语义分成三类,即不存在型空值、存在型空值和占位型(或无信息)空值。本文主要讨论未知型空值,这类信息值是存在的,但当前不知道,故称之为未知空值。

为便于讨论,下面给出本文所用到的学生选课数据库示例,其关系模式如下:

STUDENT (SNO, SNAME, AGE, SEX, DEPT) // 学生的学号、姓名、年龄、性别和系

COURSE (CNO, CNAME, PCNO) // 课程号、课程名和先修课程号

SC (SNO, CNO, GRADE) // 学生(SNO)选修某门课(CNO)的成绩(GRADE)

表1 关系 STUDENT

SNO	SNAME	AGE	SEX	DEPT
S1	李小龙	23	男	计算机
S2	张立林	NULL	NULL	计算机
S3	杨伟仁	19	男	NULL
S4	许晓萍	NULL	女	数学

^{*} 南京大学计算机软件新技术国家重点实验室开放课题基金。毛宇光 博士后,副教授,主要研究方向为数据库理论及特种数据库。徐洁磐 教授,主要研究方向为数据库系统及数据仓库。韩 波 硕士生,主要研究方向为数据库系统及理论。

wareBus. pdf

2 王光平. OO 软件总线研究. 计算机工程与应用, 2000, 3

3 Kruthoff A. Jini and Software Bus Systems. <http://www.soft-wired-inc.com/pdf/technology>

4 Jügel M L. High-level Communication in Large Simulation Frameworks. <http://www.first.gmd.de/persons/leo/pub/Soft-Bus>

对应的关系实例如表1、表2和表3所示,其中 NULL 表示相应的属性值缺失。

表2 关系 COURSE

CNO	CNAME	PCNO
C1	数据库	C3
C2	操作系统	C4
C3	数据结构	C4
C4	离散数学	NULL

表3 关系 SC

SNO	CNO	GRADE
S1	C1	92
S1	C2	NULL
S2	C1	100
S3	C3	75
S4	C1	NULL
S4	C4	65

3 三值逻辑条件表达式的求值

在单一空值(NULL)环境下,关系数据库的查询条件可能有三种情况,条件成立(T)、条件不成立(F)和不能确定条件是否成立(U),因此对查询条件的求值应采用三值逻辑。为了适应空值环境下关系数据库的应用需求,我们在文[6]中提出了一种基于命题联结词 \wedge 、 \vee 、 \neg 、 \rightarrow 和 μ (未知)的三值逻辑谓词演算系统 MF^M ,它为不完全信息数据库的理论研究奠定了逻辑基础。联结词 \wedge 、 \vee 、 \neg 和 \rightarrow 的真值表与 Kleene 三值逻辑系统相同, μ 的真值表见表4。为便于讨论,下面用例子说明在不完全信息数据库中条件表达式的求值。

例1 含命题联结词的条件表达式求值。

当 $A=3, B=4, C$ IS NULL 时,条件表达式求值结果如表5所示。

表4 μ 的真值表

A	μA
T	F
U	T
F	F

表5 条件表达式求值

条件表达式	结果	条件表达式	结果
$A=3$	T	$A>B \text{ OR } B>C$	U
$C=4$	F	$A<B \text{ OR } B<C$	T
$(B=4)\rightarrow(C=3)$	U	$\text{NOT}(A=C)$	U
$\mu(C=3)$	T	$\mu(A=B)$	F

量词有存在量词(\exists)和全称量词(\forall)两种,它们的求值方法如下:

$\exists v(p(v))$ 等价于 $F \text{ OR } p(v1) \text{ OR } p(v2) \text{ OR } p(v3) \text{ OR } p(v4) \dots$

$\forall v(p(v))$ 等价于 $T \text{ AND } p(v1) \text{ AND } p(v2) \text{ AND } p(v3) \text{ AND } p(v4) \dots$

例2 含量词的条件表达式求值。

关系 R 如表6所示,设 t 是 R 上变化的范围变量,含量词的条件表达式求值结果如表7所示。

表6 关系 R

A	B	C
1	2	3
1	2	NULL
NULL	NULL	NULL

表7 条件表达式求值

条件表达式	结果	条件表达式	结果
$\exists t(\mu(t.A>3))$	T	$\forall t(t.A>1)$	F
$\exists t(\mu(t.C=3))$	T	$\forall t(t.B>1)$	U
$\exists t(t.B>2)$	U	$\forall t(\mu(t.C>1))$	F

4 三值元组关系演算 MFMTRC

为适应不完全信息处理的需求,本节构造了一种三值元组关系演算系统 MFMTRC,它是以三值逻辑谓词演算系统 MF^M 为逻辑基础的,是对传统元组关系演算的真扩充。MFMTRC 具有较强的语言表达能力,传统的元组关系演算可看作它的特例。与传统元组关系演算的区别是,MFMTRC 合式公式中出现的联结词均为三值逻辑联结词,且增加了处理空值的联结词 μ 。

三值元组关系演算表达式的一般形式为:

$$\{t \mid \Psi(t)\}$$

其中 t 是一个元组变量,(是由原子公式、联结词和量词组成的合式公式(wff),表示所有满足 Ψ (求值结果 T)的元组 t 的集合。

下面定义三值元组关系演算 MFMTRC。

定义1 三值元组关系演算 MFMTRC 的原子公式定义如下:

(1) $R(t)$ 是原子公式,其中 R 为可含有空值的关系, t 为元组变量,表示 t 是 R 的一个元组。

(2) $t[i]\theta s[j]$ 是原子公式,其中 t, s 为元组变量, θ 为算术比较符($<, \leq, >, \geq, =, \neq$)或空值比较符 $=_{\omega}$,它表示元组 t 的第 i 个分量和元组 s 的第 j 个分量满足 θ 关系。算术比较的结果有 T、F 和 U 三种,当 $t[i]$ 或 $s[j]$ 为 NULL 时,结果为 U。对于 $=_{\omega}$,当左右两边的分量都为 NULL 时,结果为 T。

(3) $t[i]\theta C$ 或 $C\theta t[i]$ 是原子公式,其中 C 为 D_i

上的常量, θ 的含义同上。

(4) 原子公式仅由上面三种方式定义。

定义2 三值元组关系演算 MFMTTC 的公式可递归地定义如下:

(1) 每个原子公式是一个公式, 其中所有元组变量是自由变量。

(2) 若 φ 和 ψ 是公式, 则 $\varphi \wedge \psi, \varphi \vee \psi, \varphi \rightarrow \psi, \neg \psi$ 和 $\mu\psi$ 也是公式。在公式中的元组变量的自由约束性仍然和 φ, ψ 中的自由约束性一样。但可能某个元组变量 t 在一个公式中是自由的, 而在另一个公式中是约束的。

(3) 若 ψ 是公式, ψ 中有自由元组变量 t , 则 $(\forall t)(\psi)$ 和 $(\exists t)(\psi)$ 也是公式。 t 在 ψ 中是自由的, 而在公式 $(\forall t)(\psi)$ 或 $(\exists t)(\psi)$ 中变为约束的, 其它变量的自由约束性保持不变。

(4) 若 ψ 是公式, 则 (ψ) 也是公式。

(5) 公式仅由上面四种方式定义。

在 MFMTTC 中, 含有全称量词和存在量词的公式的求值如下:

$(\forall t)(\psi(t))$

=T 如果满足 $\psi(t)$ 的集合为空集或对所有的元组 t , 使得 $\psi(t) = T$;

=F 如果满足 $\psi(t)$ 的集合非空且至少存在一个元组 t , 使得 $\psi(t) = F$;

=U 否则(满足 $\psi(t)$ 的集合非空, 不存在一个元组 t , 使得 $\psi(t) = F$ 且至少存在一个元组 t' 使得 $\psi(t') \neq T$, 即 $\psi(t') = U$)。

$(\exists t)(\psi(t))$

=T 如果满足 $\psi(t)$ 的集合非空且至少存在一个元组 t , 使得 $\psi(t) = T$;

=F 如果满足 $\psi(t)$ 的集合为空集或对所有的元组 t , 使得 $\psi(t) = F$;

=U 否则(满足 $\psi(t)$ 的集合非空, 不存在一个元组 t , 使得 $\psi(t) = T$ 且至少存在一个元组 t' 使得 $\psi(t') \neq F$, 即 $\psi(t') = U$)。

公式中各种运算符的优先次序规定如下:

(1) 算术及空值比较符: $<, \leq, >, \geq, =, \neq, =_{\omega}$

(2) 量词: \forall (全称量词), \exists (存在量词)

(3) 否定词: \neg, μ

(4) 二元联接词: \wedge (合取词), \vee (析取词), \rightarrow (蕴含词)

定义3 三值元组演算表达式是形如 $\{t | \psi(t)\}$ 的表达式, 其中 t 是公式 ψ 中仅有的自由元组变量。

例3 查找“年龄小于20岁的女生”这个条件求值结果为未知的学生姓名。

$\{t | \exists (u) (\text{STUDENT}(u) \wedge t[1] = u[2] \wedge \mu(u[3] < 20 \wedge u[4] = '女'))\}$

结果为: {张立林, 许晓萍}。

例4 查找所有选修无先修课课程的学生姓名。

$\{t | \exists (u) (\exists v) (\exists w) (\text{STUDENT}(u) \wedge \text{COURSE}(v) \wedge \text{SC}(w) \wedge u[1] = w[1]$

$\wedge v[1] = w[2] \wedge v[3] = \omega \text{NULL} \wedge t[1] = u[2])\}$

结果为: {许晓萍}。

5 三值域关系演算 MFMDRC

三值域关系演算 MFMDRC 与三值元组关系演算 MFMTTC 的定义类似, 主要区别是公式中的变量不是元组变量而是取值于属性值域的域变量。三值域关系演算 MFMDRC 可在三值元组关系演算 MFMTTC 中使用下面两条替换得到。

(1) 元组变量 t 用域变量 $u_1 u_2 \dots u_n$ 替换;

(2) 第 i 个元组分量 $t[i]$ 用一个域变量 $u_i (1 \leq i \leq n)$ 替换。

三值域关系演算表达式的形式是: $\{u_1 u_2 \dots u_n | \psi(u_1, u_2, \dots, u_n)\}$

其中公式 ψ 既可以使用联结词 $\wedge, \vee, \rightarrow, \neg$ 和 μ , 也可以使用存在量词 \exists 和全称量词 \forall , 而 u_1, u_2, \dots, u_n 是 ψ 中仅有的自由域变量。

不难证明下面的定理。

定理1 三值元组关系演算 MFMTTC 和三值域关系演算 MFMDRC 的表达能力是等价的。

小结 在数据库的查询优化过程中, 常常需要将 SQL 语句转换为关系演算或关系代数等内部表示。另外, 在讨论 SQL 语言的形式语义时, 也需要一种抽象语言。本文提出了一种表达能力较强的、适用于不完全信息数据库的三值元组关系演算 MFMTTC 和三值域关系演算 MFMDRC。如何用 MFMTTC 研究 SQL 语言的形式语义, 目前已得到了一些初步结果, 许多问题还有待进一步研究。

参考文献

- 1 Codd E F. Extending the database relational model to capture more meaning. ACM Trans. Database Syst., 1979, 4(4): 397~434
- 2 Biskup J. A foundation of Codd's relational maybe-operations. ACM Trans. Database Syst., 1983, 8(4): 608~636
- 3 Yue K B. A more general model for handling missing information in relational databases using a 3-valued logic. SIGMOD Record, 1991, 20(3)
- 4 Negri M, Pelagatti G, Sbattella L. Formal semantics of SQL queries. ACM Trans. Database Syst., 1991, 17(3): 513~534
- 5 Levy A Y. Obtaining Complete Answers from Incomplete Databases. In: Proc. of the 22nd VLDB Conference, Mumbai (bombay), India, 1996
- 6 毛宇光, 徐洁磐. 不完全信息数据库的逻辑. 计算机科学, 2000, 27(10. 增): 252~254, 271