文章编号: 1006-4729(2002)01-0038-03

关系模式中候选码的求解

张 永, 顾国庆
(上海理工大学 计算机工程学院, 上海 200093)

摘 要: 关系数据库模型的应用十分广泛, 其候选码的求解已被证明是一个NP完全问题, 从闭包的角度讨论了如何求解候选码, 并对其求解过程进行了一定程度的优化. 最后提出了一些比较合理的建议.

关键词: 关系模式; 候选码; 闭包; 函数依赖
中图分类号: TP311 文献标识码: A

引 言

随着数据库技术的日益完善, 关系数据库的应用也越来越广泛, 在关系数据库理论中, 候选码(又称关键字)是一个非常重要的概念, 它能唯一标识关系的元组, 是关系模式中一组最重要的属性: 采用按码检索的方式(比如索引), 可极大地提高我们对元组的检索效率. 因此, 关系模式中的一个首要工作就是求出该关系模式的候选码.

求解一个关系模式中的所有候选码, Lucchesi和Osborn已证明是一个NP完全问题. 许多有关这方面的书中也很少讨论这一问题, 一般都是通过候选码的概念来直接求解候选码, 但实际上, 在求解过程中, 我们可以对其进行局部的优化.

1 基本知识

定义1 关系模式是对关系的描述, 它是一个五元组: $R<U, D, DOM, F>$. 由于在关系模式中, 影响数据库模式设计的主要是 U 和 F. 因此我们通常把它简记为一个三元组

$$R<U, F>$$

式中: R ——关系名;

U ——组成该关系的属性名的集合;

D ——属性组 U 中属性所来自的域;

DOM ——属性向域的映像的集合;

F ——属性间数据依赖关系的集合.

定义**2** 设 K 为关系模式 $R<U, F>$ 中的属性(组). 若 $K \xrightarrow{f} U$, 则 K 称为关系模式 R 的一个候选码(Candidate Key), 若关系模式 R 有多个候选码, 则选定其中的一个做为主码(Primary Key). 这是对候选码的一种定义, 我们也可以从闭包的角度对其进行定义.

定义**3** 设 X 为关系模式 $R<U, F>$ 中的任一属性(组), 记 $X^+ = \{A \mid X \rightarrow A$ 可根据Armstrong公理系统从 F 中导出$\}$, 则称 X^+ 为属性(组) X 关于依赖集 F 的闭包.

属性(组) X 的闭包 X^+ 可通过算法来求解.

定义**4** 设 K 为关系模式 $R<U, F>$ 中的属性(组), $K^+ = U$, 且对任一属性 $A \in K$, 有 $(K-A)^+ \neq U$, 则称 K 为 R 的一个候选码.

易知, 一个关系模式的候选码 K 具有以下两个特性:

1 唯一性 在任一给定时间, 关系 R 的任意两个不同元组, 其属性集 K 的值是不相同的. 若不然, 设有关系 $R<U, F>$, 其中 $U = K \cup W$, $X \in W$ 为关系 R 的一个非主属性; 并且在关系 R 中存在这样两个不同元组, 其属性 K 的值是相同的, 但在属性 X 上的值不同, 由函数依赖的定义可知, 属性集 K 不能通过函数决定属性 X, 从而 K 不是关系 R 的候选码.

2 最小性 属性集 U 中的任一属性都不能从集合 K 中删除掉. 若能从属性集 K 中去掉某一属性 X, 且有 $(K-X) \xrightarrow{f} U$, 显然有 $K \xrightarrow{p} U$, 从而 K 不是关系 R 的候选码.

下面给出一个简单的例子. 设有关系模式

$$R<U, F>$$

收稿日期: 2001-05-22

式中: $U=\{A, B, C, D\}$;

$$F=\{A\rightarrow B, C\rightarrow D\}.$$

易知, R 的唯一候选码是(AC). 这里我们用括号, 表示它们的属性组合才是候选码. 为了操作上的方便, 我们假设关系模式 R 上的函数依赖集 F 为最小的函数依赖集. 若 F 不是最小的, 则可通过算法来求出.

对于关系模式 R 中的属性集 U, 我们可以对其作出如下的划分: U_L 表示仅在函数依赖集中各依赖关系式左边出现的属性的集合; U_R 表示仅在函数依赖集中各依赖关系式右边出现的属性的集合; 另记 $U_B=U-U_L-U_R$, 显然, 当 $U_B\neq\Phi$(空集)时, 它表示依赖关系式左右边都出现的属性的集合. 如上例, $U_L=\{AC\}$, $U_R=\{BD\}$, $U_B=\Phi$.

2　命题的提出及其证明

命题 **1**　设有关系模式 $R<U, F>$, $X\in U$, 若 $X^+=U$, 则 X 中必定包含关系模式 R 的一个候选码.

该命题由上述的定义 4 很容易得出, 此时的 X 称为超码. 证明略.

命题 **2**　设有关系模式 $R<U, F>$, 若 U_L 非空, 则 U_L 中的任一属性必包含在关系模式 R 的候选码中.

证明: 设 K 为 R 的码, 属性 $A\in U_L$, 但 $A\notin K$. 易知, $K\rightarrow A\in F^+$, 这说明一定存在一个函数依赖: $X\rightarrow A$. 这与 $A\in U_L$ 矛盾, 从而命题得证.

命题 **3**　设有关系模式 $R<U, F>$, 若 U_R 非空, 则 U_R 中的任一个属性必定不包含在关系模式 R 的任一个候选码中.

证明: 设 K 为 R 的码, 属性 $A\in U_R$, 由定义 2 知 $K\xrightarrow{f}U$. 假设 $A\in K$, 则 $K\xrightarrow{f}U$ 可写成形如 $AW\xrightarrow{f}U$ 的形式$(A\notin W)$. 显然, 此时 $A\in U_L$, 与 $A\in U_R$ 矛盾, 故 $A\notin K$.

命题 **4**　设有关系模式 $R<U, F>$, 若 U_L 非空, 且 $U_L^+=U$, 则 U_L 为关系模式 R 的唯一的候选码.

证明: 设 K 为 R 的任一候选码. 由命题 2 可知, $U_L\subseteq K$; 此时, 可设 $K=U_LX$(其中 $X\subseteq R$, 且 $X\cap U_L=\Phi$), 即 R 的任一候选码都形如 U_LX. 又因 $U_L^+=U$, 则由命题 1 可知, $K\subseteq U_L$, 即 $U_LX\subseteq U_L$. 显然, 上式要成立, 当且仅当 $X=\Phi$, 从而 $K=U_L$, 即 U_L 为 R 的唯一的候选码. 当 U_L 为空集, 或 U_L 非空, 但 $U_L^+\neq U$ 时, 求解候选码成为一个 NP 完全的问题, 我们可以利用定义来进行求解.

3　基本算法

根据我们上面讨论的命题, 可以得出如下的求解候选码基本算法的具体步骤.

第 1 步, 求关系模式 $R<U, F>$ 的最小函数依赖集 F.

第 2 步, 按照上面的定义, 分别计算出 U_L, U_R, U_B.

第 3 步, 若 $U_L\neq\Phi$, 计算 U_L^+. 若 $U_L^+=U$, 则 U_L 为 R 的唯一的候选码, 算法结束. 若 $U_L^+\neq U$, 转第 4 步. 若 $U_L=\Phi$, 转第 5 步.

第 4 步, 将 U_L 依次与 U_B 中的属性组合, 利用上述的定义 4 判断该组合属性是否是候选码; 找出所有的候选码后, 算法结束.

第 5 步, 对 U_B 中的属性及属性组合利用上述的定义 4 依次进行判断; 找出所有的候选码后, 算法结束.

在实际的求解过程中, 对于算法中的第 4 步和第 5 步, 最多重复 2^n-1 次, 其中 n 为 U_B 中属性的个数. 若 $U_B=U$, 则为最坏情况. 我们还可以采用下述的一些基本原则来进行判断:

1　在关系模式 R 中, 若 A 为码, 且 $A\leftarrow B$, 则 B 必为码;

2　在关系模式 R 中, 若 A 不为码, 且 $A\leftarrow B$, 则 B 必不为码;

3　在关系模式 $R<U, F>$ 中, 若 K 为码, $W\subseteq U$, 且 W 中不包含 K 的任一属性, 则 KW 必不为码, 而是超码.

例: 设有关系模式

$$R<U, F>$$

式中: $U=\{A, B, C, D, E\}$;

$$F=\{A\rightarrow B, AC\rightarrow D, CD\rightarrow E, E\rightarrow C\}.$$

易知, $U_L=\{A\}$, $U_R=\{B\}$, $U_B=\{CDE\}$. 因 $A^+=\{AB\}\neq U$, 则将 A 依次与 U_B 中的属性组合, 得到:

$(AC)^+=\{ABCDE\}=U$, AC 为候选码, 从而 ACD, ACE, $ACDE$ 不为候选码;

$(AD)^+=\{ABD\}\neq U$, AD 不为候选码;

$(AE)^+=\{ABCDE\}=U$, AE 为候选码, 从而 ADE 不为候选码. 算法结束.

参考文献:

[1] C. J. 戴特著. 吴鹤龄译. 数据库系统导论[M]. 北京: 科学出版社, 1984.

[2] J. D. 厄尔曼著. 张作民译. 数据库系统原理[M]. 北京: 国防工业出版社, 1984.

[3] 李昭原, 刘又诚. 数据库系统原理与技术[M]. 北京: 北京航空航天大学出版社, 1992.

[4] 萨师煊, 王 珊. 数据库系统概论[M]. 北京: 高等教育出版社, 1991.

[5] 杨超植著. 刘动天等译. 关系数据库[M]. 北京: 电子工业出版社, 1990.

[6] 王 珊, 陈 红. 数据库系统原理教程[M]. 北京: 清华大学出版社, 1998.

Evaluation to Candidate Keys to Relational Model

ZHANG Yong, GU Guo-qing

(College of Copmputing Engineering, Uninversity of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Relational database model is used widely. It is proved that finding all candidate keys of a relational model is an NP-problem. This artical mainly discusses how to evaluate all candidate keys to a relational model by using closures, and optimizes the finding procedures to a certain extent, finally also proposes some advice on evaluating the candidate keys.

Key words: relational model; candidate key; closure; functional dependency