一个函数式数据模型FM的完备性

丁公才
(镇江船舶学院)

THE COMPLETENESS OF A FUNCTIONAL DATA MODEL FM

Ding Gongcai
(*Zhen Jiang Shipbuilding Institute*)

ABSTRACT

Model completeness is an important characteristic by which the description power of data model is measured. In this paper, a functional data model FM and query language FQL has been introduced. A new conception on model implication and completeness has been defined. According to some characteristics (such as implication, completeness, etc.) on relational model, the completeness of FM has been discussed lastly.

摘　要

模型完备性是衡量数据模型描述能力的一种重要特性。本文引入了一个函数式数据模型FM和查询语言FQL, 定义了一种新的模型蕴含和完备的概念，最后，根据关系模型的有关特性(如蕴含、完备等), 讨论FM的完备性。

§1. 引　　言

近十年之间，许多研究工作者致力于高级数据模型和程序设计语言的数据抽象研究，导致了一系列新型数据模型和查询语言的诞生。其中最引人注目的是函数式数据模型和函数式查询语言FQL。

1989年5月15日收到，1989年12月9日定稿。

所谓函数式数据模型，可以非形式的描述为一组抽象数据类型[1] 以及建立在这些类型之上的函数关系。这个概念最早由Siblay 和Kerchbery 引入[2]。FQL[3] 则完全吸收了Backus 提出的函数式程序设计语言FP 的风格[4]，它较好地解决了查询语言不能同时适用于端点用户(End-User) 和应用开发程序员的矛盾[5]，突破了传统数据模型的数据操作所依赖的冯·诺依曼风格，这使得函数模型具备了在新一代计算机上实现的可能性，因此，对函数模型的一些重要特性(如蕴含特性、完备性) 做进一步研究是很有意义的。

图1 描述了一个函数模型FM 的实例。它包括5 个类型(如EMPLOYEE 等) 和6 个函数(如ENAME 等)。每个函数用传统的数学方法描述，例如ENAME: EMPLOYEE → CHAR。

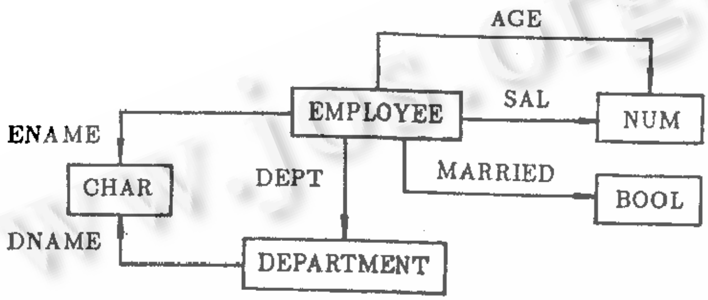

图 1

FQL 中有4 种算子，类似于FP 中的函数型的概念，它们能把简单函数构造成复杂函数。

(1) 结合算子“∘”，$f: T_1 \to T_2$, $g: T_2 \to T_3$, $f \circ g: T_1 \to T_2$。例如，DEPT ∘ DNAME 定义了这样一个函数，对于给定的一个雇员，返回他所在部门的名称。

(2) 扩展算子“$*$”，$f: T_1 \to T_2$, $*f: *T_1 \to *T_2$。$*T_1$ 表示由T_1 中所有类型值的任意子序列(流) 所构成的类型。例如，对于一个给定的雇员序列，$*$NAME 返回一个对应的姓名序列。

(3) 构造算子“[ ]”，$f_1: T \to T_1, \cdots, f_n: T \to T_n, [f_1, \cdots, f_n]: T \to [T_1, \cdots, T_n]$。这里，$[T_1, \cdots, T_n]$ 是由n 元组构成的类型。例如，对于给定的雇员，[ENAME, AGE] 返回该雇员的姓名和年龄构成的二元组。

(4) 限制算子“|”，$p: T \to BOOL, |\, p: *T \to *T$。对于$T$ 的一个序列s, $|\, p$ 产生另一个序列s', s' 是s 中所有使p 为真的类型值组成的序列。

若x 是类型值，则$\langle x \rangle$ 表示产生x 的常数函数；若T 是类型，则T 也是常值函数，它产生T 的所有类型值组成的序列。

关于函数模型和FQL 的详细描述参见[3][5]。

本文首先定义模型完备性、蕴含特性等一系列概念，然后根据关系模型的有关特性讨论函数模型FM 的完备性。

§2. 完备性定义与假设

形式上，一个数据模型实例m 是一个二元组：$m =< D, O >$，这里，D 是一组目标，它构成对现实世界客观对象的抽象描述；O 是一组算子，它能够组成对目标D 的任意操作。例如，关系模型实例$r =< D,O >$，其中D 是一组关系，O 是关系算子集合。

设m 是数据模型实例。$D(m)$ 表示m 的一组目标；$O(m)$ 表示m 的算子集合；$\theta(\Leftarrow O(m))$ 表示θ 是由$O(m)$ 作用在$D(m)$ 上完成某个操作的表达式；$\theta[y/x]$ 表示θ 中的所有x 出现用y 替换；$\theta = \theta'$ 表示θ 和θ' 的执行结果在共同的数据结构下相同。

在进行数据库设计时，人们总习惯于用一组实体和实体联系(即ER 集合) 来描述现实世界，然后再向某个数据模型转换。不严格地说，ER 集合能相当好地描述任何现实世界[6]。设e 是从ER 集合中导出的新的实体或实体联系，记作$ER \models e$。

定义2.0　任何数据模型实例的非空集合称做一个数据模型。

定义2.1　设M 是数据模型，$m \in M$，给定ER 集合。如果对任意的$e(ER \models e)$，有$\theta(\Leftarrow O(m))$，使$\theta = e$，则称m 对ER 是表达完备的。

定义2.2　设M 是数据模型，$m \in M$，给定ER 集合。如果对任意的$\theta(\Leftarrow O(m))$，有$e(ER \models e)$，使$e = \theta$，则称m 对ER 是操作完备的。

定义2.3　设M 是数据模型。M 是完备的，当且仅当对任意的ER 集合，存在$m \in M$，使m 对ER 是表达完备的和操作完备的。

在许多文献中如[6][7]，都阐述了ER 集合向关系模型转换的合理性、完整性以及算法。在[8] 中，也把关系代数作为衡量一个查询系统完备性的基准。因此，有理由做以下基本假设。

定义2.4　RM 是关系模型，当且仅当对任意的$r \in RM$，$D(r)$ 是一组关系，$O(r)$ 是关系算子集合。

基本假设：关系模型RM 是完备的。

§3. 数据模型的蕴含特性

定义3.1　设M、M' 是数据模型，$m \in M$，$m' \in M'$。如果对任意的$\theta(\Leftarrow O(m))$，存在$\theta'(\Leftarrow O(m'))$，使$\theta = \theta'$，则称$m$ 实例蕴含于m'，记作$m \subseteq m'$。

命题3.1　关系“$\subseteq$”是可传递的。

定义3.2　设M、M' 是数据模型，$m \in M$，$m' \in M'$。如果$m \subseteq m'$，$m' \subseteq m$，则称m 实例等价于m'，记作$m \Leftrightarrow m'$。

命题3.2　关系“$\Leftrightarrow$”是可传递的和对称的。

定义3.3　设M、M' 是数据模型。如果对任意的$m \in M$，存在$m' \in M'$，使$m \Leftrightarrow m'$，则称M 蕴含于M'，记作$M \subseteq^* M'$。

命题3.3　关系“$\subseteq^*$”是可传递的。

定理3.1　设M、M' 是数据模型，$m \in M, m' \in M'$，给定ER 集合。如果m 对ER 是表达完备的，且$m \subseteq m'$，则m' 对ER 是表达完备的。

定理3.2　设M、M' 是数据模型，$m \in M$，$m' \in M'$，给定ER 集合。如果m 对ER 是操作完备的，且$m' \subseteq m$，则m' 对ER 是操作完备的。

定理3.3　设M、M' 是数据模型。如果M 是完备的，且$M \subseteq^* M'$，则M' 是完备的。

§4. 关系模型RM对二元关系模型BM的蕴含性

定义4.0　BM 是二元关系模型，当且仅当对任意的$bm \in BM$，$D(bm)$ 是一组二元关系，$O(bm)$ 是关系算子集合。

设$R(X, A_1, \cdots, A_k)$ 是关系，X 是关键字，即$X \to A_1 \cdots A_k$。令$R_i(X, A_i) = \prod_{X,A_i}(R)$ $(i = 1, \cdots, k)$，称$\varsigma(R) = \{R_1(X, A_1), \cdots, R_k(X, A_k)\}$ 是R 的二元关系分解。$\bowtie_{i=1}^{k} R_i$ 记作$\varsigma(R)$。这里，$\bowtie$ 和$\prod$ 分别是自然连接和投影运算。

引理4.1　设$R(X, A, B)$ 是关系，$X \to AB$。R 的分解$\varsigma(R) = \{R_1(X, A),\ R_2(X, B)\}$ 是无损连接的。

引理4.2　若$\varsigma(R) = \{R_1, \cdots, R_k\}$ 是R 的无损连接分解，$\varsigma'(R_i) = \{S_1, \cdots, S_m\}$ 是R_i 的无损连接分解，则R 的分解$\varsigma''(R) = \{R_1, \cdots, R_{i-1}, S_1, \cdots, S_m, R_{i+1}, \cdots, R_k\}$ 是无损连接的。

推论4.1　R 的二元关系分解$\varsigma(R)$ 是无损连接的。

定义4.1　设RM 是关系模型，BM 是二元关系模型，$rm \in RM$，$D(rm) = \{R_1, \cdots, R_n\}$ 令$bm = < \bigcup_{i=1}^{n} \varsigma(R_i),\ O(rm) >$，其中$\varsigma$ 是二元关系分解，则$bm \in BM$。我们称bm 是rm 对应的二元关系模型实例，记作rm^*。

根据推论4.1 和定义4.1 可以证明，对任意的$rm \in RM$，有$rm \Leftrightarrow rm^*$。于是有以下定理。

定理4.1　$RM \subseteq^* BM$。

§5. FM 的完备性

定义5.0　FM 是函数模型，当且仅当对任意的$fm \in FM$，$D(fm)$ 是一组函数，$O(fm)$ 是FQL 的算子集合。

以下给出一个二元关系模型实例bm，向一个函数模型实例fm 转换的算法TRANS，步骤如下：

(1) 设$D(bm) = \{R_1, \cdots, R_n\}$

(2) 设ST 是fm 的类型集合，SA 是bm 的属性集合，Γ 是映射，$\Gamma : SA \to ST$。对任意的$A \in SA$，有$T \in ST$，且$T = \bigcup_{i=1}^{k} \prod_A(R_{j_i})$，其中$\{R_{j_1}, \cdots, R_{j_k}\}$ 是出现A 的关系集合。我们规定$\Gamma(\Gamma^{-1})$ 亦可把一组属性(类型) 映射成一组对应的类型(属性)。Γ^{-1} 是Γ 的逆映射。

(3) 对每个$R \in D(bm)$，设$R(X, A)$，$X \to A$，$X = \{A_1, \cdots, A_k\}$，且$\Gamma(A) = T$，$\Gamma(A_i) = T_i (i = 1, \cdots, k)$。构造：

(a) $f_0 : [T_1, \cdots, T_k] \to T$

$$f_0 = R \cup \{< t, \perp > | t \in [T_1, \cdots, T_k],\ t \notin \Pi_x(R)\}$$

(b) $f_i : [T_1, \cdots, T_k] \to T_i (i = 1, \cdots, k)$

$$f_i(t) = \begin{cases} \perp & t \notin \prod_X(R) \\ \prod_{A_i}(\sigma_{X=t}(R)), & \text{否则} \end{cases}$$

(c) 把$f_0, f_1, \cdots, f_k$ 加入$D(f_m)$。

这里，符号"$\perp$"表示函数在某处无定义值。

对$D(fm)$ 的任意操作θ, 可由以下Backus 范式FORM 产生:

$$\theta ::= \langle x \rangle \circ G | !T \circ *G \circ F \quad (1)$$

$$G ::= f | G \circ G' | [G, G'] \quad (2)$$

$$F ::= \langle 空 \rangle | F \circ | P | F \circ *G \quad (3)$$

这里，$f \in D(fm)$。设G 是FORM (2) 产生的函数，$G : T \to T'$。令$X = \Gamma^{-1}(T)$, $Y = \Gamma^{-1}(T')$, 那么G 可以等效看成一个关系$R_G(X, Y)$。于是有以下命题:

命题5.1 $R_G = \theta(\Leftarrow O(bm))$

根据FORM, 施归纳于G 的结构即可得证。

定理5.1 $BM \subseteq^* FM$

证明: 对任意的$bm \in BM$, 运用算法TRANS 把bm 转换成$fm \in FM$。以下分两步证明。

(1) $fm \subseteq bm$, 根据FORM

(a) 先证$\theta = \langle x \rangle \circ G$ 的情形。根据命题5.1, 对任意的$G : T \to T'$, 存在$\theta'(\Leftarrow O(bm))$, 使$R_G(X, Y) = \theta'$。则$\theta = \langle x \rangle \circ G = \prod_Y(\sigma_{X=x}(R_G)) = \prod_Y(\sigma_{X=x}(\theta')) = \theta''(\Leftarrow O(bm))$。

(b) 再证$\theta = !T \circ *G \circ F$ 的情形。施归纳于F 的结构:

基始: $F = \langle 空 \rangle$ 时，由命题5.1, 对于G, 有$R_G = \theta'(\Leftarrow O(bm))$, $\theta = !T \circ *G = \prod_A(R_G) = \prod_A(\theta') = \theta''(\Leftarrow O(bm))$, 这里$\Gamma^{-1}(T) = A$。

假设: $\theta_1 = !T \circ *G \circ F$ 时，存在$\theta'(\Leftarrow O(bm))$, 使$\theta_1 = \theta'$。

归纳: 若$\theta = !T \circ *G \circ F \circ | P$ 时，$\theta = \theta_1 \circ | P = \sigma_P(\theta')(\Leftarrow O(bm)$; 若$\theta = !T \circ *G \circ F * G' = \theta_1 \circ *G'$ 时，由命题5.1 可令$R_{G'} = \theta''(\Leftarrow O(bm))$, 则$\theta = \prod_Y(\theta' \bowtie R_{G'}) = \prod_Y(\theta' \bowtie \theta'')(\Leftarrow O(b_m))$。这里$G' : T_1 \to T_2$, $Y = \Gamma^{-1}(T_2)$。

(2) 给出产生$\theta(\Leftarrow O(bm))$ 的Backus 范式，运用结构归纳法容易证明$bm \subseteq fm$。

故$bm \Leftrightarrow fm$, 则$BM \subseteq^* FM$。

根据基本假设，命题3.3 和定理3.3, 有如下推论:

推论5.1 函数模型FM 是完备的。

本文并非给出函数模型的设计方法，只是借助于关系模型对函数模型做完备性讨论，进一步揭示函数模型对现实世界的描述能力。

参考文献

[1] J.M.Smith and D.C.P.Smith, Database Abstraction: Aggregation and Generalization, ACM TODS, Vol.2, No.2.

[2] E.H.Siblay and L.Kerschbery, Data Achitecture and Data Model Considerations, Proc. AFIPS National Computer Conference, Dallus, Texas, 1977.

[3] P.Buneman and R.E.Frankel, FQL—A Functional Query Language, Proc. ACM—SIGMOD Conference on Management of Data, Boston, 1979.

[4] J.Backus, Can Programming be Liberated from the Von Neumann Style? A Functional Style and Its Algerbra of Programs, C.ACM, Vol.21, No.8, 1978.
[5] P.Buneman and R.Nikkil, The Functional Data Model and Its User for Interaction with Databases, In M.L.Brodie (eds.), On Conceptional Modelling, Spring-Verlay, 1984.
[6] P.P.S.Chen, The Entity—Relationship Model Towords a Unified View of Data, ACM TODS, Vol.1, No.1.
[7] 萨师煊，王珊，数据库系统概论，高等教育出版社，1983。
[8] David Maier, The Theory of Relational Databases, Computer Science Press, Inc., 1983.

《软件学报》征稿简则

一、《软件学报》是中国科学院软件研究件主办的学术性刊物。本刊对象主要是计算机科学研究人员、工程技术人员，大专院校教师，研究生及高年级学生。

二、本刊刊登计算机理论，计算机软件的研究、设计与实现，计算机网，数据库，人工智能，计算机辅助设计及软件新技术的应用开发等方面的研究论文、短文、技术报告、研究简报、通讯及综述。

三、来稿要求和注意事项

1. 来稿内容力求正确、文字精炼，各类文稿具体要求如下：

学术论文：具有重大创新，有新的观点、见解或具有中国的特点，推动或丰富了该课题领域的研究与发展。字数不超过8000字。

短文：有创造性的或包含新的见解，但不要求是完整的工作；对已有工作的较大改进和提高。字数不超过4000字。

技术报告：有重大或较大的实用、推广价值，技术上先进和内容完整的研究报告。字数不超过8000字。

简报：技术上先进，有一定的实用或推广价值的研究成果(包括阶段成果)的扼要报道。字数不超过3000字。

通讯：有关本刊已发表文章的问题讨论或简短表达的新思想和对学科发展的建议等。

综述：针对计算机科学技术的新兴或活跃领域，对国内外发展状况及趋势的全面系统的综合评选。

2. 来稿一式三份(手写或计算机打印均可)，文章务必有二百字中文摘要和英文摘要(英文摘要放在中文摘要前面，中文题目、作者姓名及单位名称的后面。英文摘要包括题目，姓名，单位名称。英文摘要，不要另附，免得丢失)。

3. 来稿务必做到清稿定稿，一经排版，不得再作任何文字上的修改。

4. 文中的插图要精绘，图中文字书写清楚，插图在稿中所占的位置用方框标出，注明图、图注。

5. 文中的外文字母、符号必须分清大、小写，正、斜体；上、下角的字母、数码和符号，其位置高低应区别明显；容易混淆的外文字母，请在第一次出现时，用铅笔批清。文中需用黑体字之处，请在字下加波纹线。

6. 参考文献只择最主要的列入，文献按文中出现先后次序编排，书写格式如下：

期刊：[编号]作者姓名，文章题目，刊物名称，卷数：期数(年份)，页数。

图书：[编号]作者姓名，书名，出版单位，地点，年份，页数。

7. 已公开发表的文章，本刊不再予以刊登。

四、反映重大成果的稿件希附有作者单位推荐信和审查意见。

五、刊登稿件按修改最后定稿时间排序。来稿一经发表，酌致稿酬。并赠送单行本30本。不拟刊登稿件，当妥为退还。

来稿请寄北京中国科学院软件研究所《软件学报》编辑部(邮政编码100080)(北京8718信箱)