

语义数据模型 (3)

Joan Peckham Fred Maryanski

3. 模型的比较

本节用第1节提供的参数比较了第2节所描述的对象及其内容,也给出了评价这些模型的相对优点的方法。由于数据模型集合的范围或对象尚未统一,因此评价带有一定的主观性。我们仍用早先的“语义模型”定义来描述各个模型。

3.1 基于表格的比较

表1总结了所讨论的各种模型的特点。由于所选模型的多样性,表中的信息是一般性的而非具体的,更具体的参数不可能提供切实有益的比较,尤其在动态建模方面。一个由Urban和Delcambre[1986]给出的综述,为静态、动态和时态建模结构列出了详细的表格,但动态建模模型的种类却是有限的。

表1中出现了前面没有定义的一些术语,解释如下:

- (1)非结构对象表示是受限分类还是扩展分类,这依赖于系统提供的非传统数据类型的程度。
- (2)联系表示是独立的实体、表、函数或属性,都取决于模型将联系提供给用户的方式。如前所述,许多系统为语义相对化起见,为用户提供了多重联系视图。

合意的查询语言构造是下一个挑战;逻辑和物理数据库设计被认为是对当前关系系统的一个挑战,更不要说对具有更加丰富的类型系统和规则的系统了;为了对此领域中的用户提供辅助将需要数据库设计方法学和工具;规则执行的优化提出了重大的挑战,此外允许用户设计并调试面向规则的应用的工具是这一技术成功的关键。我们鼓励学术界对

(3)我们根据提供给数据库设计者的插入/删除约束将各种方法分类如下:如果提供给用户一组联系类型,其中均含有插入/删除规则,我们就认为这些是内部规则;若这些规则由系统自动支持,则可认为它们是自动的;如果用户能对各种联系类型的插入/删除语义进行选择,那么这类约束是由用户指定的。

从表中可以发现多种建模技术,例如,实体-联系模型强调网状模型,而TAXIS注重作为主要对象及操作组织特点的概括层次。这里,层次和网状性质上的细微区别对建模方法学的应用有一定的影响。层次中不允许出现循环相当于TAXIS中不允许循环继承,另外还有:

(1)几乎所有的模型都用概括和聚集作为建模抽象概念,而分类,尤其是联合,一般不被看作是基本建模概念。大多数模型趋向于建立单独的抽象概念,例如,TAXIS注重概括,而SDM则用类定义了其它性质。

(2)动态建模作为建模过程的一个不可缺少部分,令一些语义设计者感兴趣。在三个有动态建模处理能力的模型(TAXIS, SHM+和事件模型)中,有两个将一种特殊的设计方法作为似乎是模型必不可少的一部分,即模型不仅提供建模结构而且给设计者提供这些问题开展研究。

参考文献共32篇(略)

[杜小勇、萨师焯译校自《Proceedings of the IFIP DS-4 OO Databases》, July 1990]

表1 数据模型特点

	非结构 对象表示	联系表示	标准抽象 概念	层次网络	派生/继承	插入/删 除 约 束	联系语义	动态建模
E-R	受限的	独立性下表	集 合	健壮的网络		用户详细 精确定义	用户可选	
TAXIS	受限的	实体(分类)	概括、集 合、分类	健壮的层次	继 承	用户详细精 确定义除IS- A层	预定义的	事务建模 面向目标
SDM	受限的	独立性和 实体(分类)	概括、集合 分类、联合	提出通用 层次系统	精制的 可变的	自 动 的	用户定义	
FUNCTIONAL	受限的	功能性	联合、集 合、分类	对两者都无 直接支持	功能性的	用户详细 精确定义	用户定义	
RM/T	受限的	独立性	概括、集合 分类、联合	通用层次 系 统	继 承	自动的、用户 详细精确定义	预定义的	
SAM*	扩展的 (内在的专 用集合)	独立性	概括、集合 分类、联合	网 络	分类继承 的总和	自 动 的	预定义的	面向目标
EVENT	受限的	属 性	概括、分类	层次系统		自 动 的	预定义的	事务建模
SHM+	受限的	属性、实体 独立性	概括、集 合、联合	通用层次 系 统	概括的继承 联合、层次	内 在 的	预定义的	事务建模

某种建模活动顺序以定义数据库模型。

(3) 插入/删除约束条件的定义是多样的。一些模型(SAM*, SDM)把这些约束作为联系定义的一部分,其它的模型(函数式数据模型、实体-联系模型)允许用户定义自己所需的规则,而另一些模型(RM/T)则允许内部约束和用户自定义约束的混合出现。

(4) 派生的支持方法也很多,许多模型把通过概括联系的继承性作为唯一的派生形式,而如SDM这类模型则有多种不同的派生。

(5) 只有最新的或最专门化的模型为信息建模提供了丰富的数据类型,但不能简单地转化为ASCII码。如表中所示,这些数据模型提供了一些非常好的例子,但在这一方向的工作仍在继续[Copeland and Maier 1984; Woelk等1986]。

每种模型的建模观点都不相同,这一点可以从前面各节的描述中清楚地看出,表1试图进一步强调这种区别。尽管大多数模型致力于解决概念建模中的普遍问题,但仍有许多模型是用于特殊环境的,从而假定这些

特性最适合这些环境。为了进一步评价这些模型,必须确定评价的目标,这可以从下述两方面进行:

(1) 从相同性及差异性出发的模型分类法,或

(2) 针对特殊环境选择最佳模型。

在实践中,设计者将面临后一种决定。我们为划分不同语义模型提出了各种参数,并讨论了对给定的任务,选择一种“最佳模型”的方法。这就是本文的实质所在。

3.2 评价语义模型

尽管在语义数据模型中有各种概念建模方法,但所有模型的作者似乎都一致认为,其主要目标是促进数据库的建模和使用。多数设计者认为,下述内容有助于实现此目标:

(1) 模型是否能独立应用,或是象SAM*只针对特殊的环境。

(2) 联系是否能高效地开发带有所有内部语义(插入/删除约束,最基本的约束,等等)的程序包,就象SDM和SAM*一样,或是否数据库设计者应能明确指定每一联系的语义。实体-联系模型提供了后一种选择,

因为对每一种联系，设计者都指明诸如维数和插入/删除约束等性质。

(3) 联系的结构是复杂的还是基本的。例如，二元模型 [Abrial 1974, Azmoodeh 等1986, Bachman 1983, Bracci等1976] 是所有联系由基本二元联系组成的数据模型。Brodie [1984] 将其归类为不可约模型，因为信息在模型内是原子表示的，而不是复杂的事实组合；Tsichritzis 和 Lochovsky [1982] 将二元模型归类为面向图的模型，因为联系经常用图来表示（关系模型为面向表的初等模型）。为了说明基本二元联系有可能表示复杂的联系，可以考虑构成实体的聚集联系。例如，设实体BOOK由属性Title, Page-length, Author（指向另一实体）和Publisher（指向另一实体）组成，可以认为此联系是BOOK及其四个属性间的一个1:4联系。它将BOOK对象与来自四个二元联系：BOOK-Title, BOOK-Page-length等的每一属性的一次出现联系起来。虽然用二元模型有可能构成复杂联系，但可认为提供给用户的是最基本的模型。这种方法与诸如SDM之类提供内部的更高级联系的模型是有区别的。

(4) 联系是否能与概念级上的实体相区别。在一些模型，如E-R模型中，联系具有与实体不同的语义，是基本建模元素；另一方面，在函数式模型中，实体与联系均表示一定的函数。

(5) 要强调哪些抽象概念。某些模型把一、两个抽象概念作为其主要建模工具，例如，SDM大量使用了分类和聚集，而TAXIS则注重概括层次，其它一些方法，如SHM+等，也支持了大量的抽象概念。提供少量基本建模概念的模型的开发者们相信，少许有力的自由选择将使建模过程更为简单。

(6) 动态建模应遵循什么方法。SAM* 提供了抽象数据类型，而TAXIS中为概括层次，事件模型中为信息流。

上述原则可作为用于评价语义数据模型的特征。如果评价的目的就是为分类确定模

型间的共性和差异，则这些特性可构成表1中的列。使用这方面的信息，就可望象Brodie [1984] 那样将模型分组，开发一个如图20所示的连续体。

表1中的参数可用以为特定的任务选择模型。另外，考虑下列三条原则 [March 等1984]，它们可以量度个别设计者将模型用于特定问题领域的简易性。

- (1) 用户对所提出的建模结构的意义（或语义）的理解能力。如果用户较难理解或错误理解建模工具，则语义模型的优越性就完全丧失了。
- (2) 查询表示的简易性。语义模型发展的动力之一就是解决在物理及数据结构一级上操纵数据库的艰巨性。
- (3) 建模结构语义的易说明和易维护性。无论模型提供的是一组简单的插入/删除约束集合或是复杂的约束定义工具，其目的应是相同的：语义应是易定义且自动维护的。

一种被广泛应用的评价策略是在表中的每一列附加权，如表2中，权指明了每一问题领域中每一规则的相对重要性。表中的每一行所包含的分数表明相对于每一列的未来模型的价值。表3提出了适合于特殊应用环境的假想模型M1和M2的评价，M1的总计分数较高，反映了评价者们偏爱其描述联系的方法及它“更易于理解”。

在这些评价过程中，主观性是固有的，

关系式RM/T	
函数式	→ 增强的完整性
E-R	→ 联系作为函数
SDM	→ 明确的联系
TAXIS	→ 实体类
事件	→ 概括
SHM+	→ 动态建模
SAM*	→ 联合与分类
	→ 专用类型

图20 数据模型连续体

表2 以权表示的特点

	非结构化 对象	关系表示	标准抽象 概念	网络层次	派生/继承	插入/删除 约束	关系语义	动态建模
权	0	25	15	10	15	10	20	5

表3 数据模型比较

	非结构化 对象	关系表示	标准抽象 概念	网络层次	派生/继承	插入/删除 约束	关系语义	动态建模	总计
权	0	25	15	10	15	10	20	5	
M1		10	7	5	8	7	9	3	790
M2		6	8	5	6	7	4	3	600

然而,回想语义数据模型的目标,其中之一便是事实表示的结果应十分接近于用户的理解,因此,对一给定的事实,用户的理解在给定企业模式的模型评价中是重要的。多个评价者参与评价过程,可在一定程度上减少主观性。

4 我们向何处去?

尽管研究者们较早认识到了表示抽象概念和支持联系的重要性,但在语义数据建模领域还存在一些其它问题。探索模型的对象、操作和联系语义的各种意义仍是重要的。有些想法可能符合前面列出的要求,从而产生了可简化数据库设计任务的概念模型。下面给出了这种观点,它们将继续受到现代数据模型研究者的关注。

SAM^{*}提出了适用于统计学应用和 CAD/CAM 的数据类型来支持非传统数据的观点,其它研究者则认为提出一些基本原理以支持特殊类型的设计更为有效 [Copeland and Maier 1984]。这样就导致了一种更灵活的模型——其目标不是特殊应用而是特殊类型的复杂化表示,文本、声音(包括语音)、数字图象和复杂图形向量的建模原语被用于处理多介质信息 [Christodoulakis 等 1986, Woelk 等 1986], 这些多介质信息综合利用了语义数据建模原理和面向对象的语言思想以

支持形形色色的和非标准化的数据库应用。

另两个实质上进展不大的方面是动态和时态建模技术(意指包括显式时间和事件概念)。所有的语义系统都支持静态建模技术,不过,尽管研究者们已讨论了时态建模 [Snodgrass and Ahn 1986] 和动态建模的重要性,但也只能找出少数模型包含这些功能。Urban and Delcambre [1986] 调查了五个模型——RM/T、TAXIS、SDM、SHM+ 和事件模型,发现大部分模型提供动态技术,也有一些提供了有限的时态建模性能。表 1 中含时态建模信息的一栏的确较缺乏内容。尽管如此,若要概念建模工具提高后能以一种完备的方法描述企业模式,则这种描述就必须能表示时态的性态和事件。

已提出了各种时间表示法。Ariav [1986] 定义了面向时态的数据模型 TODM, 它以立方体形式把时态数据高效地表示在关系数据模型中, Ariav 的工作将关系操作扩展到了时态数据。Snodgrass and Ahn [1985, 1986] 亦将关系模型作为表示时态数据的起点。他们区分了事务时间和有效时间,前者是指信息存入数据库的时间,而后者则是信息有用的那一段时间。使用这两种概念,他们定义了静态关系和历史关系。Shoshani and Kawagoe [1986] 的方法与其他人不同,他们采用了时间序列,即时间值排序作为基本的组织概念,他们还讨论了时态数据的物理组织。

Shiel[1983]提出了以时间概念作为基本抽象概念的时态层次数据模型 (THM), 其工作详细地描述了时态对概括和特化语义方面的影响。Castilho等 [1982]开发了一种在数据模式规格说明中处理时态语义的语言。时态数据研究中的一个关键概念是时态约束, 它允许用户在约束条件中将时间因素和数据值都包含在内 [Ariav 1986, Kung 1984]。

新的信息处理领域的概念建模的需求主要推动了语义数据模型的发展, 本文讨论了几种数据模型, 能满足大部分企业模式的建模要求。提高数据建模能力所需的另一种方法是自动产生数据模型以适合特殊应用环境的要求。数据模型编译器 [Maryanski等1986, 1987]和EXODUS [Carey等1986, Richardson and Carey 1987]研究了这种方法。在可变数据模型生成器方面, 关键的问题是概念数据模型语义的形式说明及表示, 第一步应确定表示模型的实体、联系、操作和约束等抽象概念。在数据模型编译器中用于表示数据模型语义的语言出现在Hung和 Maryanski [1988]的著作中, EXODUS项目开发了E程序设计语言 [Richardson and Carey 1987]作为基本的说明机制。如果数据模型产生器的想法是可行的, 那么就有可能修改语义数据模型来满足企业模式的要求。这样, 设计者头脑中的模型与数据库系统中的表示之间的差距就缩小了。

(接第48页)

的主体并非全部相信其它主体所传递来的信念和知识, 笔者在一般意识逻辑模型基础上曾提出一个信念依赖模型^[16], 用于研究和描述多主体环境下的信念和知识的传播特性。

4. 共同知识与逻辑万能问题

如前所述, 共同知识与逻辑万能问题是关于知识的推理这个领域中最重要两个基

作为此处所提出的模型说明, 语义数据建模仍处于研究阶段, 商业市场中将不会超出关系模型, 直到性能合理的语义数据模型出现为止。按年代计, 语义数据模型比关系模型晚七年才出现, 而年轻的产品已开始涌现, 成熟的系统可望在几年内产生。这些新产品的市场目标不是商业数据处理而是科学与工程产生的大量数据的管理。商业市场上语义数据模型的影响可能受下述条件限制:

- (1) 商业市场的惰性——除非利益远大于开销, 否则把大量数据转换到新模型的过程就不会出现。
- (2) 在科学/工程/生产市场的机会——对高级数据模型的需求已经存在。

对概念建模未来的一个合理预测将影响程序设计语言的发展。一些语言 (如FORTRAN和COBOL) 自出现以来就一直被使用者青睐, 同时新的语言不断出现, 许多语言实际上影响了研究者和应用程序员。设计者总是为一个任务寻求最佳模型, 随着应用复杂性的持续上升, 设计者对概念建模的要求亦相应增强, 且新的模型将不断产生。本文中讨论的这些模型是有代表性的下一代模型, 但不是最后一代。

(全文完)

参考文献 [略]

[纪岳、何未艾、王红、赖维生、陆强、钟颖译自、姚卿达校自ACM Computing Surveys, 1988, vol.20, №3]

本问题。至今为止, 已提出了许多不同的解决方案。但仍未取得理想的结果。许多计算机科学工作者和逻辑工作者都在为之而努力。相信在不久的将来, 认识逻辑及其与知识的推理的有关技术将会得到更为广泛的应用。这也可以视为思维科学在计算机科学中的应用的一个起点。

(参考文献略)