

数据模型的智能化扩充研究*

姚卿达 纪岳

(计算机软件研究所)

摘 要

本文就如何在关系数据模型中,引入人工智能、知识工程的成果提出了看法。工作目标是提高关系模型的概念抽象能力及引进演绎推理机制,文中给出了智能化数据库的一种体系结构以及关系模型智能化扩充的一种方法。

关键词 关系数据模型,人工智能,知识工程,概念抽象,演绎推理,智能化扩充

1 问题的提出

数据库理论和方法已经受到广泛的关注和应用,成为当前信息资源管理的最重要技术之一。但传统的数据库只能存储数据与数据之间的联系信息和进行事实检索。它在提供更友好的用户界面,提供更丰富、更有力的反映现实世界信息的能力,以及运用事实进行咨询和决策方面都显得不足。能否让数据库存储知识,并进行演绎推理,这是数据模型智能化扩充所要解决的重要问题,也是当今数据库系统发展的一个新方向^[1]。因此,数据库(DB)、人工智能(AI)和程序设计语言(PL)相结合^[2]已成为数据库发展的必然趋势,也是计算机科学发展的主流之一。

本文从概念抽象及演绎推理两个方面对“数据模型的智能化扩充”这一主题进行论述,并提出了研究和实现的方案。目标是将ES与DBMS技术相结合,产生通用化、智能化的DBMS,从而使通常只存储数据(事实)和只进行事实检索的DB,具有存储知识(书本、常识、经验)和演绎推理能力。

2 概念抽象与演绎推理

2.1 概念抽象与框架系统

框架系统是一种知识表示语言,它来源于人工智能中的语义网络。上述二者均是基于一个认识论上的假定,即:人的知识能够组织成一个相关概念的网络。框架系统可以支持诸如: Is-a, Superclass, Ako, Subset 之类的概念抽象技术。而概念抽象本身固有的继承特性^[3]使得框架系统比那些基于关系、记录或指针的系统具有更强的知识表达能力。继承特性由简单的广度优先搜索方法加以实现。框架系统能够支持面向对象的程序设计及复杂信息系统设计,而且允许用户定义新的关系或抽象功能,如:

本文1988年6月16日收到

* 国家自然科学基金资助项目

aggregation, elaboration, abstraction, revision, default, precedence, similarity, 和 causality 等。在实际的数据库设计中,还能对聚集关系(aggregation)和因果联系(causality)做出进一步的语义区分。

在数据模型中引入概念抽象机制,用框架表示描述的数据库系统具有分类与层次的性质。这将表现在以下5个方面:

①分类是一个大型信息系统的组织原则。继承特性能够使我们以清晰、易懂的方式构造一个系统。把不同数据类型的共同性质(属性)集中于一处,有助于减少冗余和避免更新异常。

②继承特性也支持了系统的可扩充性。运用它可以方便地在现行系统中引入新的数据类型和联系,实现系统自身的完善和扩充;又能够把现行的系统引入到未来的、规模更大的系统中去。

③对于表示为分类、层次的数据库系统,只要加入新的子数据类型,就能实现系统的扩充,而且可以在不影响现行数据类型的前提下,对系统的功能进行更新或扩充。

④可以化简数据库的维护。因为对数据类型的更新是没有重复的,而且在更新的起始时候,不必考虑每个类型的细节,详细的说明可以逐步地展开。

⑤分类与层次可以帮助用户更好地理解数据库的涵意,构造出合理的查询视图,并提出有意义的查询访问。

2.2 演绎推理与产生式系统

产生式系统已在大多数专家系统的实现中得到了应用,特别适于表示演绎性的知识,它作为一种控制陈述语言和过程性知识的机制在应用中证实是很有用的。产生式系统由一组逻辑规则组成。规则具有Horn子句的形式,清晰明了,而且规则之间的相互作用非常小,易于对大量知识的系统进行修改。通过对规则集合的结构进行分析,可以大大提高产生式系统的演绎推理效率,并提高用户对跟踪求解问题的透明度。

在数据模型中引入产生式规则,把数据库划分成外延数据库(由表示事实的关系组成)和内涵数据库(由表示约束及演绎过程的规则组成)两个部分,从而使数据库具有演绎推理的功能,其主要表现在以下6个方面:

①强有力的演绎功能。能够利用内涵数据库中的规则,通过计算和搜索外延数据库中的事实,导出丰富的结论。

②减少数据冗余。在传统关系数据库中,数据之间大量存在的各种逻辑联系(如函数依赖、多值依赖等),可以用规则来表示和存贮,从而节省空间,提高搜索效率。

③用户友好的接口。规则可以用来定义关系,为用户定义数据库视图提供很好的灵活性,而且用户查询实际上是定义一个新关系,而这个关系可以直接表示为一个内涵关系,使查询相当简洁,而且具有递归的功能。

④解决数据的一致性。在数据库应用中,库的内容可能经常随时间变化,由于数据之间复杂的依赖关系,个别数据的变化可能会引起许多数据的变化。用规则描述数据一致性,就能方便地解决此问题。

⑤数据库模式的自动生成。如果某一内涵关系使用频繁,为提高效率,可以借助推理机构将其转化为外延关系。

⑥可扩展性好。通过增加知识表示、知识获取、不精确推理和解释功能，可以开发更高级的知识处理系统^[4]。

3 智能化数据库

本节从概念抽象与演绎推理两条线索来讨论关系数据库的智能化扩充问题，并在综合前人工作的基础上，提出自己的看法。

3.1 概念抽象技术与关系数据库

最先明确地把概念抽象引入数据库系统的是Smith, J. M和S. C. P. Smith. 他们在[5]中提出了运用抽象机制：聚集(aggregation)和概括(generalization)来设计、组织及实现数据库的原则和方法。此后，概念抽象技术一直是被作为数据库设计方法来使用的，它也启发人们从抽象的角度拓展数据模型的语义表达能力^[6]。接着人们开始从理论上探讨概念抽象技术与关系数据库模型相结合的可能性。在[7]中，Aliteboul, S 揭示了在聚集作用下，数据库基本运算（如：联接、投影、选择）的一组性质，以及数据库依赖（FD和JD）与数据库模式的有关性质。

在关系数据库理论的范围内，人们也开始了对概念抽象技术的研究。Abiteboul, S在[8]中这样写到：“包含依赖(Inclusion Dependency, 简称IncD)是ISA关系的一个自然结论，从而一出现就引起人们极大的兴趣。E/R模型可以用FD和IncD自然地描述。而且IFO模型(有特化(specialization)但没有概括(generalization))也可以用FD和IncD来描述。”已经有一批论文比较深入地研究了IncD的性质以及与FD的联系^[9]等。

在关系数据库与抽象机制相结合的系统实现方面，也开展了初步的工作。在[10]中，C, Beerl等人通过扩展数据库查询语言QUEL，从而实现了比较高级的数据库语义查询功能。

3.2 能演绎推理的关系数据库

自从逻辑程序设计语言 Prolog 问世以来，人们就开始考虑把能进行大规模数据(事实)处理的DBMS与具有运用逻辑规则进行演绎推理的 Prolog 系统耦合起来，并以此为核心构造演绎数据库系统。国内外许多研究人员遵循这条思路开展了许多工作^[11]。

为此，[12]专门提出了一个研究数据库的新数学理论——扩充模型论，它以模型论为基础，能表示数据库的各种查询、操纵功能，又具有演绎功能。这一理论不仅反映了整个数据库的面貌，还能作为研究数据库的数学工具。

另外，N, Spyrotas在[13]中提出了分区语义模型(Partition Semantic Model)，它可以看作由关系数据模型扩充得到的演绎数据库模型。这个模型以代数格的理论作基础，从而得到关于关系属性、数据约束、数据库模式的统一描述。这一方面简化了关系模型的语法表示，又扩充关系模型的语义表达及演绎推理能力。

分区语义模型为我们提供了更丰富的数据约束，它运用陈述性、过程性规则以及相应的演绎控制机构，能够以“是”、“否”、“不知道”等方式回答用户的提问。而且该模型的有关性质与关系数据库研究的最新进展联系十分密切。

3.3 智能化数据库及其体系结构

纵观上述对于语义数据模型以及演绎数据库的研究,“关系”仍然是构造数据库的基本单位。表现有二:其一是“事实”将作为“关系”存储在数据库中,另外有部分规则亦能表示为关系。这意味着:现行的关系数据库理论与技术,将对数据模型智能化扩充的工作起重要作用,而且既然关系数据模型是以谓词逻辑为基础的^[14],所以我们认为:以谓词逻辑为基础,就能扩充关系模型的语义描述及演绎推理能力。

对于智能化数据库必须从丰富的语义表示能力与强大的演绎推理能力两个方面去理解,才能比较全面。这不仅更完整地吸收、综合了数据库系统的最新成果,而且为下一步知识库研究与设计奠定了更广阔的基础。在我们的智能化数据库系统中,由陈述性、过程性规则以及演绎推理规则组成的叫内涵动态数据库(也叫规则库)。动态的含义主要指:能够组织、更改库中的规则,并运用它们推演出新的知识,由根据概念抽象机制组织的事实(关系)以及相应的继承规则组成的叫外延静态数据库(也叫事实库)。这里静态应作这样的理解:继承规则表示的是不同抽象层上概念之间的必然联系,它数量少,内容单纯,而且一经确定就不须更改了。图1给出了智能化数据库系统的系统结构,主要强调数据与规则之间的相互联系。

智能化数据库必须具备语义概念抽象表示能力和逻辑演绎推理能力这两个方面的功能。尽管前一种功能也可以用后一种方法和机制来实现,但是实现二者之间的区分,不论是对更好地理解智能化数据库,还是以关系模型为基础来扩充且高效率地实现智能化数据库都意义重大。我们在传统演绎数据库体系的基础上,把规则库一分为二:静态规则意指表示概念抽象的继承规则,用来对事实进行分层与组织;

动态规则也就是通常所说的演绎规则与推理规则,为了提高系统效率,充分利用现行关系DBMS的功能,我们用关系的形式表示和处理规则。当用户输入一个查询时,演绎机构首先对该查询进行分析,进行事实检索与推理演绎。然后演绎机构运用演绎与推理规则,对演绎路线进行优化,确定涉及的事实关系。再以此为线索访问事实数据库,给出用户所需的查询结果。

4 关系数据模型的智能化扩充

我们区分理论与实现两个方面对这个问题进行研究,并介绍在实验性系统

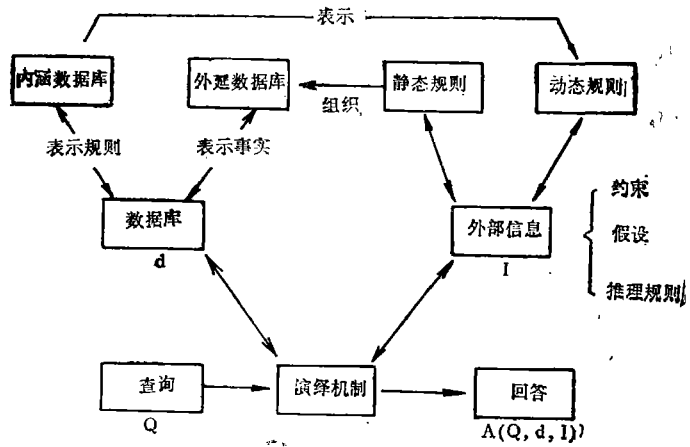


图1 智能化数据库的体系结构

Fig. 1 The architecture for intelligent database systems

RCEAS^[15]中所采用的方法。

4.1 系统实现方面

4.1.1 扩充A 演绎规则与知识表示。这里主要考虑如何在数据库定义语言(DDL)中增加描述概念抽象机制、演绎推理法则的功能。

4.1.2 扩充B 演绎推理策略——深度优先搜索(depth--first search)。这里必须考虑正向推理(由因到果)和反向推理(由果到因),其中都需要建立推理网,找出因果(果因)路径,并处理相关的关系属性。

4.1.3 扩充C 数据库查询。除了常规的利用关键字或内容进行事实检索之外,我们希望对SQL语言进行扩充,增加因果检索,回答How? Why?之类的提问等。

4.1.4 扩充D 触发操纵。数据间的语义联系使数据一致性、完整性的维护更加困难,数据的更新会引发一系列相关数据的更新操作,为此必须存贮、查找触发路径。

在RCEAS中,仍采用关系模式来表示规则模式,用关系表示规则。用于表示规则的关系模式叫CF模式,它有四种类型的属性:原因属性、结果属性、联系属性、操作属性。形式为:

CF(Cause, Effect, Connected, Pattern)。CF模式与基于规则的知识表示法一致,可以解释为

```
IF      cause 1
        cause 2...
THEN   effect 1
        effect 2...
```

RELATIONSHIP ATTRIBUTE

A1 [of...]

A2 [of...]

PATTERN (skeleton of a sentence explaining the meaning of causation rules)

其中Cause(cause1, ...)为原因关系名, Effect(effect1, ...)为结果关系名。通过逻辑推理以及关系运算,可将多因多果规则特化为单因单果来处理,所以文中只考虑单因单果的CF模式。

在RCEAS中,触发操作路径通过关系模式CE来存贮,CE有5个属性:关系1,关系2,追加关系1时对关系2的操作,更计关系1时对关系2的操作,删关系1时对关系2的操作,其中关系1和关系2是触发操作联系对,不一定具有因果联系。RCEAS解决了前述扩充A—D的一系列问题,在通常DBMS功能基础上增加了规则管理、触发操作管理以及演绎推理等机构,由于充分利用了关系理论和方法,所以是一种较为自然、效率较高的扩充方案。

4.2 理论研究方面

4.2.1 研究A 演绎规则与数据库依赖均能表示为一阶谓词的形式,而且对于数据依赖(如FD, MVD等)与命题逻辑、Horn子句、知识库规则之间的关系已有了一些成果。这方面的深入研究对于大面积移植数据库技术和原理到知识库研究中有较大意义^[16]。

4.2.2 研究B 概念抽象技术与关系理论结合,已引起广泛的注意。需要在概念抽象的环境中(如:在关系之间存在聚集,概括等语义联系时)研究FD, MVD以及数据库模式与关系运算的性质。

4.2.3 研究C 关于演绎数据库模式的研究有待开展。需要研究FD在推演时反映出来的性质,进而研究能够取得较高的演绎与维护,更新效率的规则结构,最后导出最优的智能化数据库模式。

5 结束语

在数据模型的智能化扩充之外,还可以对其进行横向的扩充:分布式数据模型;纵向扩充:时态历史数据模型。这方面的工作已初步展开。这是很有前途的研究方向。

下面用图2和图3来总结我们前面的研究思路及其内容。

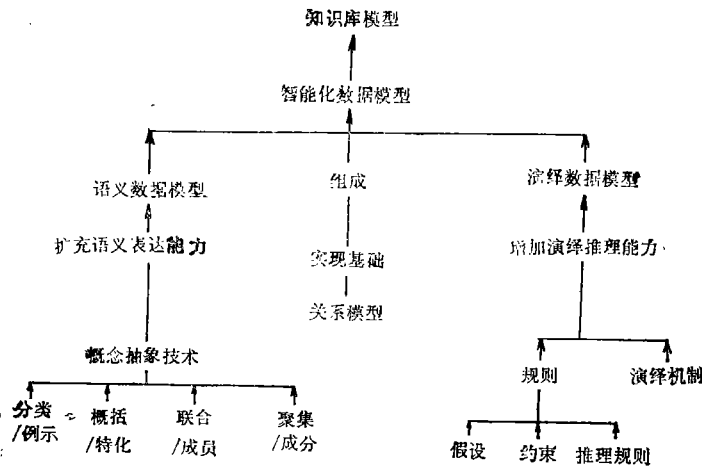


图2 数据模型智能化扩充思路

Fig. 2 The clue of intelligent extension of database model

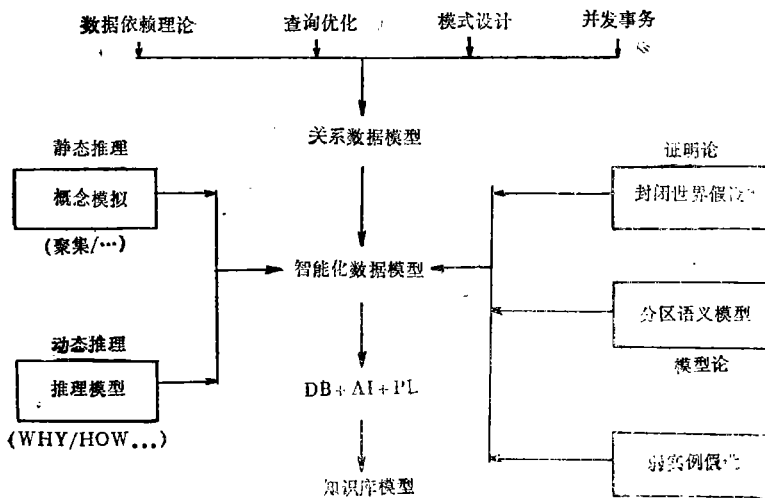


图3 研究计划及其内容

Fig. 3 Research plan and its content

参 考 文 献

- [1] Gad Ariav et al., *New Directions for Data base Systems*, Ablex Publishing Corp, 1986
- [2] Larry Kerschberg, *Expert Database Systems Proceedings From the 1st International Workshop*, The Benjamin/Cummings Publishing Company, Inc., 1986
- [3] 刘梦赤等, 概念模拟方法, 计算机科学, 1987, 5
- [4] 吴泉源等, 软件产业, 1988, 1
- [5] Smith J M et al., *ACM TODS*, 2 (1977), 2, 105~133
- [6] Su Y W et al., *A Semantic Association Model for Conceptual Database Design*, E-R Approach to System Analysis and Design, North Holland, 1980
- [7] Abiteboul S et al., *IFO: A Formal Semantic Database Model* PODS, 1984
- [8] Serge Abiteboul, *JACM*, 12(1985), 1
- [9] Casanova M. A et al., *Inclusion Dependencies and Their Interaction with Functional Dependecies*, Proceedings of the ACM Symposium on Principles of Database Systems, March, 1982, 171~176
- [10] Beeri C et al., *Compatible Attribute in a Universal Relation*, Proceedings of the ACM Symposium on Principles of Database Systems, 1982, 55~62
- [11] David Maier, *Databases in the Fifth Generation Project: Is Prolog a Database Language?*, New Directions for Database Systems Ablex publishing Corp., 1986
- [12] 徐洁磐, 计算机技术, 1987, 5
- [13] Nicolas Spyrotas, *ACM TODS*, 12 (1987), 1
- [14] Gallaire H et al., *ACM Computer. Surv* 16, 2 (1984), 153~185
- [15] Yao Q D et al., *Relational Data Model with Cause-Effect Association*, 12th, VLDB, 1986
- [16] Marek W, *Completeness and Consistency in Knowledge Base Systems*, 1st International Conference on Expert Database Systems, 1986

Research on Intelligent Extension of Database Model

Yao Qingda* Ji Yue

Abstract

The intelligent extension of data model is one of the active fields of current research in databases. This paper brings out authors' own ideas on how to introduce the achievements of Artificial Intelligence and Knowledge Engineering into the relational data model. The goal of our work is to improve the conceptual abstract abilities of relational model and to introduce the deductive inference mechanisms. An architecture of the intelligent databases and an intelligent extension method of the relational model are given.

Keywords relational database model, artificial intelligence (AI), knowledge engineering (KE), conceptual abstraction, deductive inference, intelligent extension

* Department of Computer Science