

时态数据库

可视化

数据模型

(19)

计算机科学1999Vol. 26No. 12

数据库

69-72 时态数据库的可视化方法与技术*)

Visualization Methods and Technique for Temporal Databases

方思行 陆颖

(暨南大学计算机科学系 广州 510632)

TP311.13

Abstract In this paper, a conceptual temporal model, which combines qualitative and quantitative temporal constraints, is presented. In the model an event (temporal object) no longer corresponds to a single temporal interval (a temporal interval only represents an occurrence of the event). It is a set of intervals meeting the given constraints conditions. By adding specific temporal fields into a relational database and defining a temporal view, each temporal object can be transformed into a visualized entity in the coordinate plane and a visualized temporal constraint propagation algorithm is introduced. This visualization methods and technique is helpful for understanding and analyzing a temporal network and detecting a consistent and robust solution in the network.

Keywords Temporal database, Qualitative constraints, Quantitative constraints, Temporal reasoning, Visualization

1 引言

时态数据库是将时间的概念结合到传统的数据库中。近十年来,人们对此所作的许多研究主要集中在把关系数据模型扩展为时态模型以及(或)把关系查询语言 SQL 和 Quel 扩展为时态的模式。除了强调关系模型之外,也有一些研究是将 ER 和其他的语义数据模型扩展到时态方面。时态数据模型的形式化和时态数据库的物理数据组织结构也受到了人们的关注。

时态知识的表示和时态推理的问题一直是时态数据库研究的热点。近几年,人们提出了几种基于约束的模型来进行时态推理。在这些模型中,时态推理被形式化为约束满足问题。模型中,变量是时态对象,诸如时间点和时间区间;而时态关系则形式化为对时态对象在时间轴上所处位置的约束。关于时态数据模型,最简单的一种做法就是在存储的数据对象上附加一个时间区间,但是这种方法所能表示的时态知识十分有限,目前影响较大的时态数据模型可以大致分为定性的和定量的两种,前者强调时序关系的描述,而后者则强调事件发生情况的描述。在定性模型中,变量是时间区间或时间点,而约束是定性的。在定量模型中,变量是时间点,而约束是定量的。这两类模型已经被结合成一种新

模型^[1-4]。我们在本文中提出的模型建立在 Meiri 模型^[4]的基础上,在该模型中变量可以是点或者区间,约束则有三种类型:点与点之间的定量约束以及点与区间和区间与区间之间的定性约束。

回答约束处理中的查询可以归结到确定网络的一致性、计算一个约束方案和计算最小网络的任务。当时间用整数表示时,决定一致性的问题是 NP 完全的^[2,4],对定性网络,计算最小网络是 NP 困难的^[2],无论是在定性或定量模型中,这种计算复杂性都源于变量之间的析取关系,并且发生在许多应用问题里。

因为回答这类查询本质上是难处理的,所以我们将在这篇文章中将集中研究一种方法,以便高效率且有效地决定一致性和进行时态推理。

我们知道,图形表示可以大大加强人们对复杂关系与结构的理解能力。信息的图形表示在人类的交流和认知过程中一直发挥着重要的作用。人从图形中获取其信息要比通过阅读和理解文字来获取信息要快得多,因为人能够认识图形中元素的空间位置的意义,并且迅速地找出它们的相互联系。所以,采用可视化计算机辅助技术,通过丰富的图形学技术对各类信息建立特殊的映射关系,给以动态的显示,各类信息的操作、访问在图示上给予体现,有可能为时态数据库与时态

*) 广东省自然科学基金资助项目。方思行 教授, 副教授, 从事人工智能(主要是非单调推理、时态数据库方面)、计算机数值模拟、科学计算可视化的教学和科研工作。

推理提供一种新的解决方案,研究知识的可视表示、可视推理和数据空间到可视空间的转换已成为人工智能领域中极为重要的发展方向。基于这种思想,本文提出了一个新的时态概念模型,它不仅集成了定性与定量约束,而且有利于实现时态对象的二维表示。对于在新模型中如何利用可视化进行时态推理,本文也作了一些讨论。

2 可视化方法与技术

对时态数据库采用可视化方法,可以大致分为两类:

(1)静态可视 主要针对信息的可视化。例如在数据结构层进行可视化,形成可视结构。对于对象,在实体层形成可视实体,建立一种内在的映射关系,不仅反映信息本身的对应,而且可以把图形操作与对象检索更新对应起来。

(2)动态可视 主要针对流程、操作、过程。对时态数据库的查询操作通过直接图形操作反映在时态视图中。对时态数据库的推理过程可以采用各种可视策略,例如在时态视图上进行相关时态对象的提取、剪裁等。另外还有对调试、维护过程的可视(主要体现在工具的界面上)。

可视化技术主要体现在,在图形窗口环境中把各类窗口实体(window, frame, panel, icon, cursor, scrollbar等)同知识实体、可视操作建立映射关系。根据用户的习惯和接受能力来显示、生成图形信息,决定定位和布局,提供各类可视设备(坐标、网格、颜色)来组织各类信息。

由上可知,时态数据库的可视化工作主要体现在两个方面,即基于可视化技术的实体可视化和基于可视实体的动态过程可视化。

3 时态对象的可视化

3.1 时态表示语言

可视知识实体的表达是整个可视化的关键部分,而用特征来描述对象是一种有效和实用的方法。为此我们给出一种基于定性和定量约束的时态表示语言。

首先,时态对象定义为时间点或时间区间。时态对象 O_i 和 O_j 之间的定性约束表达为析取形式: $(O_i r_1 O_j) \cup (O_i r_2 O_j) \cup \dots \cup (O_i r_k O_j)$, 即 O_i 和 O_j 之间的定性约束 $R = \{r_1, r_2, \dots, r_k\}$ 。其中 $r_i (i=1, 2, \dots, k)$ 是两个时态对象间可能存在的基本关系^[1], 这种基本关系有以下两种类型:

(1)区间与区间之间的基本关系表达为关系集 $\{b, m, s, d, f, o, bi, mi, si, di, fi, or, =\}$ 。另外,我们将与时态对象对应的时间点看作是延续时间为零的区间,则

时间点与区间的关系也被看作区间与区间之间的关系。

(2)区间端点与该区间以及区间与该区间端点之间的基本关系,表达为关系集 $\{s, f\}$ 和 $\{si, fi\}$ 。

针对时间点的绝对位置或时间点之间的距离,定量约束有两种类型:

(1)一元约束,作用在点 P_i 的一元约束限定时间点 P_i 的位置在一给定时间区间集: $(P_i \in I_1) \cup \dots \cup (P_i \in I_k)$ 。

(2)二元约束,点 P_i 和 P_j 之间的二元约束限定时间点 P_i 和 P_j 之间的距离 $P_j - P_i$ 的可能值: $(P_j - P_i \in I_1) \cup \dots \cup (P_j - P_i \in I_k)$ 。

在这两种情况下,定量约束都表达为一个区间集 $\{I_1, \dots, I_k\}$, 其中每一区间既可以是开区间、闭区间、也可以是半开半闭区间。

3.2 一个基于网络的时态模型

在我们的模型中,时态对象可以是时间点或时间区间;时态对象间的约束可以是定性或是定量的,定性约束表达为关系集合 R , 定量约束表达为区间集 I 。这样,有关时态对象之间的时序知识便构成一个时间关系网络,而时态对象是这个网络中的节点,两个节点间的弧表示相应的时态对象之间的时序关系集。具体而言,时间区间表示事件,对应于网络中的时间区间变量 I , 该变量 $I = [I_1, I_2]$ 与它的端点时间点变量 I_1, I_2 之间存在定性约束,即 $I_1 \{s\} I$ (s 表示 starts), $I_2 \{f\} I$ (f 表示 finishes)。对时间点变量 I_1 或 I_2 , 根据有关信息,给出一元定量约束 $\{[ES, LS]\}$ 和 $\{[EF, LF]\}$, 即 $\{[最早开始时间, 最晚开始时间]\}$ 和 $\{[最早结束时间, 最晚结束时间]\}$, 而 I_1 与 I_2 之间给出二元定量约束 $\{[d, D]\}$, 即 $\{[事件所需最短时间, 事件所需最长时间]\}$ 。这样,任何一个事件的发生情况就可以用一个六元矢 $\langle ES, LS, EF, LF, d, D \rangle$ 定量地表达清楚;事件之间的时序关系则通过给出时间区间变量间的定性约束 R 来表达。

需要指出的是,在我们的模型中,事件(时态对象)已不再是与一个单一的时间区间相对应(一个时间区间只表示事件的一次发生),而是满足给定约束条件的时间区间的集合。

定义1 时态对象 O_1 和 O_2 之间存在定性约束 r , 即 $O_1 r O_2$ 当且仅当存在正整数 i 与 j , 使得 $o_i \in O_1, o_j \in O_2$ 并且 $o_i r o_j$, o_i 和 o_j 分别叫做 O_1 和 O_2 的一次发生。

例如, W, L 分别表示“小王去上课”和“小李去上课”这两个事件。小王在7:20到7:30之间离开家去上课,路上要花20到30分钟。李小明在7:35到7:45之间离开家去上课,路上要花15到20分钟。如果已知“小王7:20出发,7:40到校”和“小李7:42出发,7:58到校”,则 $W \{< L$ 。

3.3 时态视图

我们通过在一般的关系数据库中增加一些时态字段的办法将其扩充为时态数据库,具体方法是:

Name	ES	LS	EF	LF	d	D	Constraints	...
------	----	----	----	----	---	---	-------------	-----

其中,Name:事件名称;ES:最早开始时间;LS:最晚开始时间;EF:最早结束时间;LF:最晚结束时间;d:事件所需最短时间;D:事件所需最长时间;Constraints:指向一个记录所有事件与相关事件间的定性时态关系的文件,其数据结构为 $[O_1, O_2][r_1, r_2, \dots, r_n]$.

由于每个时态对象都对应着一个六元矢 $\langle ES, LS, EF, LF, d, D \rangle$,利用这些时态字段,可以定义一个时态视图,从而每个时态对象可以转化为平面上的一个可视实体,它是时态对象的二维表示.可视实体中的每个点代表着满足上述六元矢的事件的一次发生.由于时态视图是一个基本表的行列子集视图,所以它是可更新的,即可执行UPDATE、DELETE和INSERT三类更新操作.可以证明,上述模型的表达能力与Itay Meiri提出的扩充的定性网络相当^[1].但不同的是,利用该模型有利于实现时态对象的可视化.一个可视实体的一般形式如图1所示.

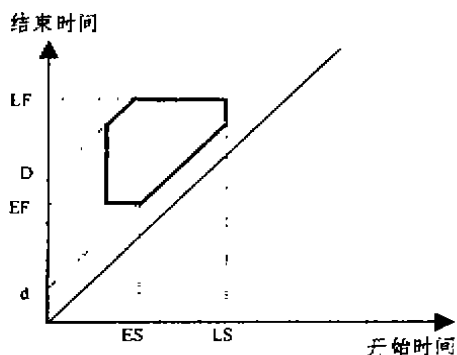


图1 一个可视实体的一般形式

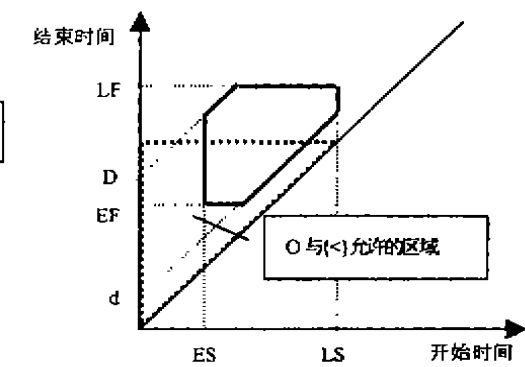
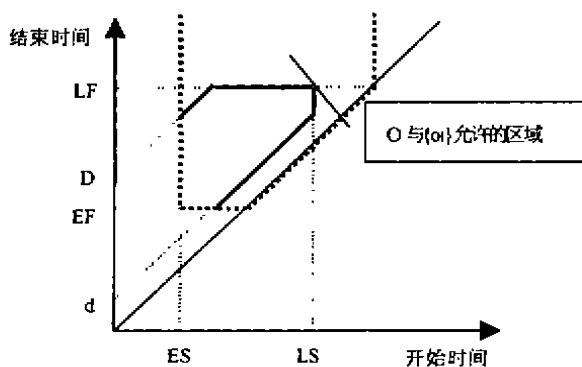


图3 可视实体与一个基本定性时态关系允许的区域

4 可视化时态推理

4.1 时态视图中定性约束和定量约束之间的关系

通过时态视图,我们容易发现时态对象的定性约束和定量约束之间存在着重要的联系.

定义2 令 r 是两个时间区间之间存在的一个基本定性时态关系,即13种可能的基本关系: $\{b, m, s, d, f, o, bi, mi, si, di, fi, oi, =\}$ 中的某一个; e 是某事件的一次发生,与 e 具有关系 r 的所有事件的发生的集合 $\{x|x(r)e\}$ 在坐标平面上确定了一个区域,称为 e 和 r 允许的区域.

图2是 e 和13种可能的基本关系所允许的区域分布.

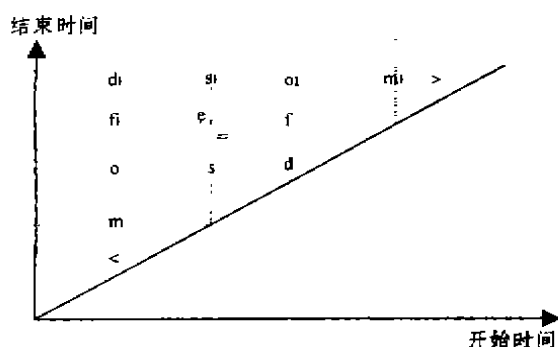


图2 事件的一次发生 e 和13种可能的基本关系所允许的区域分布

定义3 令 r 是两个时态对象之间存在的一个基本定性时态关系,则时态对象 O 与 r 所允许的区域是 O 的每次发生 o 与 r 所允许的区域 $\delta(o, r)$ 的并集,记为 $\delta(O, r)$.

例如,图3表示了事件 O 分别与 $\{oi\}$ 和 $\{<\}$ 所允许的两个区域.

4.2 可视化的时态约束传递

我们已经知道定性网络存在如下的层次关系^[1]。

$\text{net}(\text{CPA}) \subset \text{net}(\text{PA}) \subset \text{net}(\text{IPA}) \subset \text{net}(\text{IA}) = \text{QN}$

其中: CPA—Convex Point Algebra; PA—Point Algebra; IPA—Interval-Point Algebra; IA—Interval Algebra; QN—Qualitative Network。

$\text{net}(\text{CPA})$ 、 $\text{net}(\text{PA})$ 、 $\text{net}(\text{IPA})$ 、 $\text{net}(\text{IA})$ 分别表示可用 CPA、PA、IPA、IA 表示的定性网络。沿着这一层次关系,从 $\text{net}(\text{CPA})$ 到 $\text{net}(\text{IA})$,我们可获得更强的表达能力,但同时网络逐渐丧失了易处理性。判断 PA 网络一致性的复杂度为 $O(n^2)$,而判断 IPA 或 IA 网络的一致性为 NP 完全的。

因此,我们可以断定,不存在判断本文提出的一般性网络一致性的多项式算法。针对这种情况,我们首先应用一种基于三角形的定性约束传递算法^[6],它利用时态网络中的道路相容性对时间约束关系进行筛选,其算法的时间复杂度是 $O(n^3)$,但是这种方法并不能保证网络的整体相容性。

根据定义3,我们容易看出,如果两个时态对象之间存在着某种定性的时态约束,那么利用它们所允许的区域便可以进行时态对象的裁剪。这实际上是一种弧相容性的检查,它将有助于判断网络一致性问题的解决。

现在假设有 n 个事件 $O_1, O_2, \dots, O_n$, 事件 O_i 和 O_j 之间存在定性时态约束 R_{ij} , 数据集 Queue 用来记录在约束传递过程中时态视图发生变化的时态对象。在算法开始前,预先将 n 个事件放在 Queue 中,可视化时态约束传递算法如下。

```
while Queue is not empty do
  Get next  $O_i$  from Queue;
  for  $j=1, n$  do
     $\theta \leftarrow \emptyset$ ;
    for each  $r \in R_{ij}$  do
      if  $O_i \cap \delta(O_j, r) = \emptyset$ 
        then  $R_{ij} \leftarrow R_{ij} - \{r\}$ ;
      else  $\theta \leftarrow (O_i \cap \delta(O_j, r)) \cup \theta$ 
      endif
    endfor
    if  $\theta = \emptyset$  then exit {signal contradiction};
    elseif  $\theta \neq O_i$  then
       $O_j \leftarrow \theta$ ; Queue  $\leftarrow$  Queue  $\cup \{O_j\}$ ;
```

(上接第78页)

也是解决问题的途径之一。总之,数据挖掘算法与数据库间的接口问题仍有很多工作要做。

参考文献

- 1 Frawley W J, et al. Knowledge Discovery in Databases: An Overview. In: G. Piatetsky-Shapiro & W. J. Frawley, eds. Knowledge Discovery in Databases. Menlo Park,

```
endif
endfor
endwhile
```

需要注意的是,当两个时态对象之间的定性时态约束是多个基本时态关系的析取时,应用上述算法将可能导致在可视时态实体中出现若干个“洞”,即被删除的区域,这在进一步的约束传递中可能会大大增加算法的复杂性。因此,在实际应用中,应该采取一些办法使得两个时态对象之间的定性时态约束只含一个基本时态关系,有关的讨论请参阅文[5,7]。

结论 本文给出了一种集成定性与时态约束的可视时态概念模型。在模型中,事件(时态对象)不再与一个单一的时间区间相对应(一个时间区间只表示事件的一次发生),而是满足给定约束条件的时间区间的集合。通过在关系数据库中增加特定的时态字段,并且定义一个时态视图,使得每个时态对象可以转化为平面上的一个可视实体,从而导出了可视化的时态约束传递算法。在理论上,该算法并不能保证时态网络的一致性,它只是一种筛选算法。如果我们不仅希望找到一个一致的解,而且希望解是“健壮的”(不会因小小的扰动而失效),那么本文提出的时态推理的可视化技术将有助于找出这样的解。

参考文献

- 1 Allen J F. Maintaining knowledge about temporal intervals. Communication of the ACM, 1983, 26(11): 832~843
- 2 Dechter R, et al. Temporal constraint satisfaction problems. Artificial Intelligence, 1991, 49: 61~95
- 3 Kautz H, Ladkin P. Integrating metric and qualitative temporal reasoning. In: Proc. AAAI-91, 1991, 241~246
- 4 Meiri I. Combining qualitative and quantitative constraints in temporal reasoning. In: Proc. AAAI-91, 1991, 260~267
- 5 Schwalb E, Dechter R. Processing disjunctions in temporal constraint networks. Artificial Intelligence, 1997, 93: 29~61
- 6 方思行. 时序推理系统 TRS 的设计与实现. 模式识别与人工智能, 1991, 4(2): 11~18
- 7 方思行. 一种有效的 R-时刻表综合算法. 华南理工大学学报, 1995, 23(9): 43~48

CA, AAAI Press, 1991

- 2 Quinlan D. C4.5: Programs for Machine Learning. Morgan Kaufmann, 1993
- 3 John G H, Langley P. Estimating continuous distributions in Bayesian classifiers. In: P. Besnard & S. Hands, eds. Eleventh Annual Conf. on Uncertainty in Artificial Intelligence, Morgan Kaufmann Publishers, San Mateo, 1995
- 4 Cover T M, Thomas J A. Elements of Information Theory. John Wiley & Sons, 1990