

基于 XQuery 符号表示法的模糊时空数据查询

柏禄一¹, 严丽², 马宗民¹

(1. 东北大学 信息科学与工程学院, 辽宁 沈阳 110819; 2. 东北大学 软件学院, 辽宁 沈阳 110819)

摘 要: 讨论了模糊时空数据查询的概念, 研究了模糊时空数据时态查询、空间查询、属性查询和时空查询的各类查询形式及查询特性, 基于 XQuery 提出了模糊时空数据查询的统一符号表示法, 并对典型查询进行了举例说明。此外, 还讨论了 XQuery 模糊时空扩展问题, 包括 XQuery 表达上的扩展、处理上的扩展及体系结构上的扩展。提出的表示法可以表示时空查询语言 STQL 和移动目标查询语言 FTL 中规定的各类查询, 并可以消除由不同查询表现形式带来的语义混淆, 以对所有模糊时空数据查询进行统一表示。

关 键 词: 模糊时空数据; 查询语言; XQuery; FLWOR 语法; 符号表示法

中图分类号: TP 302.1 **文献标志码:** A **文章编号:** 1005-3026(2013)04-0505-04

Fuzzy Spatiotemporal Data Queries Based on XQuery Symbolic Notation

BAI Lu-yi¹, YAN Li², MA Zong-min¹

(1. School of Information Science & Engineering, Northeastern University, Shenyang 110819, China; 2. School of Software, Northeastern University, Shenyang 110819, China. Corresponding author: BAI Lu-yi, E-mail: blovdew@126.com)

Abstract: The conception of fuzzy spatiotemporal data queries was discussed. Fuzzy temporal queries, spatial queries, attribute queries, and spatiotemporal queries of fuzzy spatiotemporal data were studied. A symbolic notation for fuzzy spatiotemporal data queries was proposed on the basis of XQuery. The typical queries were shown by examples and explanations. In addition, the fuzzy spatiotemporal extensions of XQuery were discussed, including extensions of representation, extensions of management, and extensions of architecture. Queries in STQL and FTL could be represented using the proposed symbolic notation. What's more, semantic confusion caused by different forms of spatiotemporal query representation could be eliminated, and a uniform representation of various types of queries on fuzzy spatiotemporal data could be obtained.

Key words: fuzzy spatiotemporal data; query language; XQuery; FLWOR syntax; symbolic notation

时空数据库作为数据库领域的一个重要分支,受到了越来越多的关注^[1]。另一方面,由于测量误差、观测误差、数字化误差等,模糊性普遍存在于时空数据库应用系统中。时空数据库查询语言是查询和操作时空数据库的接口,在模糊时空数据查询方面,主要是对经典的查询语言进行扩充。

1) 基于 SQL 的时空数据库语言。基于 SQL 对经典时空数据库语言进行扩充包括对 SQL 的

子句不作任何扩充和对 SQL 的子句进行扩充^[2]。在此基础上,可以利用 FSQL^[3]中对模糊特性的一些扩充规则对模糊时空数据库语言进行扩充。

2) 基于 OQL 的时空数据库语言。一些学者提出了以 ODMG 对象模型为基础^[4]对时空数据库语言 OQL 进行扩展。在此基础上,也可以利用 FOQL^[5]中对模糊特性的一些扩充规则对模糊时空数据库语言进行扩充。

收稿日期: 2012-07-12

基金项目: 国家自然科学基金资助项目(60873010, 61073139); 教育部新世纪优秀人才支持计划项目(NCET-05-0288); 中央高校基本科研业务费专项资金资助项目(N090504005)。

作者简介: 柏禄一(1984-),男,辽宁丹东人,东北大学博士研究生; 马宗民(1965-),男,黑龙江齐齐哈尔人,东北大学教授,博士生导师。

3) 基于XML的时空数据库语言. 根据模糊时空数据的表示模型^[6]以及Fuzzy XQuery^[7]中对经典XQuery的扩展实现对模糊时空数据查询的扩展.

在模糊时空数据查询方面, 由于需要考虑传统关系查询中的各显式属性以及时间维和空间维中的各隐式属性, 加之模糊时空数据内各属性互相依赖以及模糊时空数据的连续性等特点, 已有的模糊时空数据查询扩展中同一种查询在不同的模糊扩展方法中表现为不同的形式. 本文将基于XML的查询语言XQuery的一种符号表示法给出一种模糊时空数据的查询方法, 从语义上对各类时空数据查询的模糊扩展统一了表示, 结合模糊时空谓词, 可用于对模糊时空数据查询作进一步的研究.

1 模糊时空数据查询概念

模糊时空数据查询: 允许被查询的模糊时空数据在时间、空间、属性以及时空方面存在模糊性的查询. 在时间方面, 可以是在时间点和时间区间上存在模糊性; 在空间方面, 可以是在位置上存在模糊性; 在属性方面, 可以是属性本身或属性数量上存在模糊性; 在时空方面, 可以是通过时空操作之后存在模糊性.

本文将基于XML的查询语言XQuery给出模糊时空数据查询的一种符号表示法, XQuery使用的是FLWOR表达式表达查询, 因此首先讨论模糊时空数据在FLWOR表达式中的模糊性.

在FLWOR表达式中, 除了“order”子句其余子句都有可能是模糊的. “for”子句的模糊性体现在循环变量的数量上, “let”子句的模糊性体现在绑定变量的数量上, “where”表达式的模糊性体现在条件上的模糊, 可以是时间条件上的模糊、空间条件上的模糊、属性条件上的模糊以及时空操作上的模糊, “return”表达式的模糊性体现在返回结果上的模糊, 包括返回变量数量上的模糊以及数值上的模糊.

2 模糊时空数据查询的符号表示

本节给出模糊时空数据查询基于XQuery的符号表示法. 为简单起见, 空间方面只考虑二维空间对象; 时态方面可以是用户自定义的时间.

模糊时空数据基于XQuery的符号表示定义如下:

Spatiotemporal query: = “for” qualifier (“let” qualifier) * “where” spatial qualifier “and” temporal qualifier “and” attribute qualifier “return” qualifier “” qualifier

Spatial qualifier: = x qualifier “” y qualifier

x qualifier: = qualifier_x y qualifier: = qualifier_y

temporal qualifier: = valid time “” transaction time

valid time: = qualifier_v transaction time: = qualifier_t

attribute qualifier: = qualifier_a

qualifier: = C (确定值) | F (模糊值) | S (单值)

| R (范围值) | E (范围集合) | * (任意值) | OP (谓词操作)

根据给出的符号表示法, for子句建立了一个绑定循环; let子句是将值赋给变量, let的数量根据需要可以是零个或多个; where子句从3个方面对结果进行过滤: 空间方面、时态方面及属性方面. 在空间方面包括横纵坐标, 在时态方面包括有效时间和事务时间; return子句返回查询结果, 包含查询的结果是确定值还是模糊值, 以及一个表达式. 在给出的符号表示法中, 符号“and”用来分隔各个限定语入口, 符号“”用来分隔限定语内部的各个部分.

< qualifier > 的取值定义为7种: 确定值、模糊值、单值、范围值、范围集合、任意值、谓词操作. 为了区分XQuery中的各个限定部分, 对各限定部分的< qualifier >加上下标予以区分. 其中下标 x 和 y 分别表示空间限定部分中的横纵坐标; 下标 v 和 t 分别表示时态限定部分中的有效时间和事务时间; 下标 a 表示属性限定部分中的属性.

从另一方面来看, 根据给出的符号表示法, 模糊时空数据查询分为显式限定查询与隐式限定查询. 其中, 显式限定查询可以直接通过限定语的显式属性得到, 而隐式限定查询则需要对查询对象通过时态谓词、空间谓词或时空谓词等操作实现. FLWOR表达式中各子句中的限定语可以根据< qualifier >的不同取值对已有查询进行分类. OP (谓词操作) 包括时态谓词、空间谓词、时空谓词, 可以是一元或者二元谓词操作, 主要用于对时间、空间、属性等限定语的隐式限定. 时态谓词、空间谓词、时空谓词的内容, 可参见文献[8].

3 模糊时空数据查询

3.1 模糊时空数据时态显式限定查询

模糊时空数据时态显式限定查询是直接通过时态限定语的显式属性便可以得到的查询. 表示为

Spatiotemporal query: = “for” qualifier (“let” qualifier) * “where” $\text{qualifier}_x \wedge \text{qualifier}_y$ “and” $\text{qualifier}_v \wedge \text{qualifier}_t$ “and” qualifier_a “return” qualifier “” qualifier

其中黑体部分 qualifier_v 和 qualifier_t 是时态显式限定语, 可以直接通过其显式属性得到。

例1 返回所有目标颜色为黑色时的当前可能位置。

表示法: for * let S let S where $S_v \wedge S_t$ and S_a return $F \wedge *_{x \text{ and } *_{y}}$

说明 “所有目标” 表示没有对查询对象作任何限定, 可以是任意值, 因此表示为*。

例2 返回1号目标颜色为黑色时在时间 $[t_2, t_4]$ 和 $[t_7, t_8]$ 时间段的可能历史轨迹。

表示法: for S let S let S where $E_v \wedge S_t$ and S_a return $F \wedge *_{x \text{ and } *_{y}}$

说明: 有效时间的限定为两个时间段的集合, 因此表示为 E 。

3.2 模糊时空数据空间显式限定查询

模糊时空数据空间显式限定查询是直接通过空间限定语的显式属性便可以得到的查询。表示为

Spatiotemporal query: = “for” qualifier (“let” qualifier) * “where” $\text{qualifier}_x \wedge \text{qualifier}_y$ “and” $\text{qualifier}_v \wedge \text{qualifier}_t$ “and” qualifier_a “return” qualifier “” qualifier

其中黑体部分 qualifier_x 和 qualifier_y 是空间显式限定语, 可以直接通过其显式属性得到。

例3 返回曾经可能位于沈阳中心(41°47'44"N, 123°26'53"E) 位置上的所有黑色目标。

表示法: for * let S where $S_x \wedge S_y$ and $*_v \wedge S_t$ and S_a return $F \wedge *$

说明 “曾经可能” 没有对有效时间作限定, 因此表示为*。

例4 返回当前沈阳市内各目标为黑色时的可能位置。

表示法: for * let S let S where $R_x \wedge R_y$ and $S_v \wedge S_t$ and S_a return $F \wedge *_{x \text{ and } *_{y}}$

说明 “各目标” 没有对查询对象作限定, 因此表示为*。

3.3 模糊时空数据属性显式限定查询

模糊时空数据属性显式限定查询是直接通过属性限定语的显式属性便可以得到的查询。表示为

Spatiotemporal query: = “for” qualifier (“let” qualifier) * “where” $\text{qualifier}_x \wedge \text{qualifier}_y$ “and” $\text{qualifier}_v \wedge \text{qualifier}_t$ “and” qualifier_a “return” qualifier “” qualifier

其中黑体部分 qualifier_a 是属性显式限定语, 可以直接通过其显式属性得到。

例5 返回1号目标当前可能的位置。

表示法: for S let S let S where $S_v \wedge S_t$ and $*_a$ return $F \wedge *_{x \text{ and } *_{y}}$

说明: 本查询为模糊属性显式限定查询, 由于没有对属性作任何限定, 因此属性表示为*。

例6 返回与1号目标颜色相同的所有目标。

表示法: for S let S where $*_x \wedge *_{y \text{ and } *_{v \wedge *_{t \text{ and } OP_a}}$ return $F \wedge *$

说明: 由于没有对位置和时间作限定, 因此表示为*; “与1号目标颜色相同” 蕴含的信息为需要对各对象的颜色属性与1号目标的颜色属性进行比较操作 Equal。Equal 为属性操作。

3.4 模糊时空数据时空隐式限定查询

模糊时空数据时空隐式限定查询是需要通过时态谓词、空间谓词或时空谓词对查询对象进行操作才能得到的查询。表示为

Spatiotemporal query: = “for” qualifier (“let” qualifier) * “where” $OP_x \wedge OP_y$ “and” $OP_v \wedge OP_t$ “and” qualifier_a “return” qualifier “” qualifier

其中涉及到时空元素 OP_x, OP_y, OP_v, OP_t 是需要通过时态谓词、空间谓词或时空谓词对查询对象进行操作才能得到的查询, 其中 OP_x, OP_y, OP_v, OP_t 可以是一个或多个。

例7 返回在沈阳市内逗留时间与1号目标相同的各黑色目标。

表示法: for * let S where $R_x \wedge R_y$ and $OP_v \wedge S_t$ and S_a return $F \wedge *$

说明: 本查询为隐式查询, “逗留时间与1号目标相同” 蕴含的信息为需要对各对象有效时间与1号目标的有效时间作 Equal 操作, Equal 为时态谓词。

例8 返回1号目标在沈阳市内时可能威胁到的各黑色目标。

表示法: for * let S where $OP_x \wedge OP_y$ and $OP_v \wedge S_t$ and S_a return $F \wedge *$

说明: 该查询是隐式查询, 蕴含的信息为需要对各对象的运动轨迹与1号目标的运动轨迹作时空操作 stOverlap。stOverlap 为时空谓词。

4 XQuery 的模糊时空扩展

在传统 XQuery 中引入模糊时空信息之后, 需要对其进行模糊时空扩展。

在 XQuery 的表达上, 也就是 FLWOR 表达

式上的模糊时空扩展在前几节内容中已有讨论.

在 XQuery 的处理上,需要在原有查询体系结构的基础上增加模糊时空处理部分来处理模糊时空信息.模糊时空处理部分包括模糊时空句法分析器、模糊时空语义分析器、模糊时空代码生成器、模糊时空翻译器及模糊时空处理器等等.

在 XQuery 查询语言的体系结构上,需要增加模糊时空操作部分、模糊时空执行树部分及相应的缓存部分.

5 与相关工作的比较

与模糊时空数据查询相关的研究主要有时态数据库查询语言、经典时空数据库查询语言以及移动目标数据库查询语言.本文提出的基于 XQuery 符号表示法的模糊时空数据查询还可以表示时空查询语言 STQL 和移动目标查询语言 FTL 中规定的各类查询,并可以根据限定部分的不同对其进行归类.例如:

(a) STQL 查询: Select $e \rho$

From Each Targets $e \rho$

Where o .targeted = "1" And e .color = "黑色" And distance (at (e .route , ρ pos) at (o .route , ρ pos)) < 5 000 And PR_inside (ρ pos ,Region (沈阳市))

该查询用符号表示法可表示为: for * let S where $OP_x \wedge OP_y$ and $OP_v \wedge *_{\rho}$ and S_a return $C \wedge *$.

(b) FTL 查询: Retrieve e

From Targets

Where end_time (Intersect (e .threatenarea ,object (1).position)) < now And begin_time (Intersect (e .threatenarea ,object (1).position)) And begin_time (INSIDE (e .Region (沈阳市)) And INSIDE (object (1) ,Region (沈阳市)) And e .color = "黑色"

该查询用符号表示法可表示为: for * let S where $OP_x \wedge OP_y$ and $OP_v \wedge *_{\rho}$ and S_a return $C \wedge *$.

(c) FTL 查询: Retrieve e

From Targets

Where end_time (Intersect (e .threatenarea ,object (1).position) And INSIDE (e .Region (沈阳市))) < 1 200 And begin_time (Intersect (e .threatenarea ,object (1).position) And INSIDE (e .Region (沈阳市))) > 1 000 And e .color = "黑色"

该查询用符号表示法可表示为: for * let S

where $OP_x \wedge OP_y$ and $R_v \wedge *_{\rho}$ and S_a return $C \wedge *$.

查询(a)和查询(b)虽然对某些限定语的操作不同(包括位置信息的横纵坐标以及有效时间的操作),但同属一类.查询(b)和查询(c)虽然在形式上相似,但查询(b)中有效时间需要通过操作得到,而查询(c)中有效时间是一个范围值,在本文提出的符号表示法中类别明显不同.

6 结 语

由于 XML 自身的优点有利于时空数据,特别是模糊时空数据的表示和查询,本文讨论了模糊时空数据查询概念,研究了模糊时空数据时态查询、空间查询、属性查询和时空查询的各类查询形式以及查询特性,基于 XQuery 符号表示法提出了一种模糊时空数据查询方法,并对典型查询进行了举例说明.此外,从 3 个方面讨论了 XQuery 的模糊时空扩展问题.基于本文所提表示法可以消除由不同查询表现形式带来的语义混淆,以达到对各类模糊时空数据查询进行统一表示的目的.未来的研究工作主要有:多维空间查询语义的扩充;嵌套查询语义的扩充;XQuery 模糊时空扩展的具体细节问题,包括体系结构以及操作等.

参考文献:

- [1] Abraham T ,Roddick J F. Survey of spatio-temporal databases [J]. *Geoinformatica* ,1999 ,3(1): 61 - 99.
- [2] Kim D H ,Ryu K H ,Kim H S. A spatiotemporal database model and query language [J]. *The Journal of Systems and Software* ,2000 ,55(2): 129 - 149.
- [3] Galindo J. New characteristics in FSQL: a fuzzy SQL for fuzzy databases [J]. *Computer Journal of WSEAS Transactions on Information Science and Applications* ,2005 ,2(2): 161 - 169.
- [4] Kuroki S , Lshizuka K , Makinouchi A. Towards a spatiotemporal OQL for four dimensional spatial database system hawks [C]//The 8th International Workshop on DEXA. Toulouse ,1997: 142 - 147.
- [5] Nepal S ,Melbourne R. A fuzzy object query language (FOQL) for image databases [C]//The 6th International Conference on Database Systems for Advanced Applications. Hsinchu ,1999: 117 - 124.
- [6] Stefanakis E. A unified framework for fuzzy spatiotemporal representation and reasoning [C]//The 20th International Cartographic Conference. Beijing ,2001: 2678 - 2687.
- [7] Goncalves M ,Tineo L. Fuzzy XQuery [C]//Soft Computing in XML Data Management. Berlin: Springer ,2010: 133 - 163.
- [8] Erwig M ,Schneider M. Spatiotemporal predicates [J]. *IEEE Transactions on Knowledge and Data Engineering* ,2002 ,14(4): 881 - 901.