

基于 XML 的不确定时空数据模型

柏禄一, 徐长明, 刘旭龙, 于长永

(东北大学 信息科学与工程学院, 辽宁 沈阳 110819)

摘 要: 由于在传送过程中存在数据失真、丢失或者网络延迟等原因, 时空数据的不确定性普遍存在且无法避免, 因此需要研究一个折中办法, 使其不影响对数据的处理. 基于 XML 建立不确定时空数据模型表达时空数据具有更多优势, 可以更好地管理不确定时空数据. 传统的关系数据库对数据的表示以关系为基础, 对时间的表示有大量的数据冗余, 使其不能很好地表示不确定时空数据. 基于 XML 构建不确定时空数据的表示模型, 对各个节点分类, 对于特殊的不确定节点, 引入概率属性来表述, 提高了数据表示能力.

关 键 词: 数据库; 时空数据; XML; 不确定; 数据模型

中图分类号: TP 311

文献标志码: A

文章编号: 1005-3026(2016)11-1546-05

An Uncertain Spatiotemporal Data Model Based on XML

BAI Lu-yi, XU Chang-ming, LIU Xu-long, YU Chang-yong

(School of Information Science & Engineering, Northeastern University, Shenyang 110819, China. Corresponding author: BAI Lu-yi, E-mail: baily@neuq.edu.cn)

Abstract: Uncertainty in spatiotemporal data is common and inevitable because of data distortion, data loss, or network delay. Accordingly, an approach that can process spatiotemporal data without being affected should be proposed. XML makes it powerful to model, represent, and manage uncertain spatiotemporal data. Data in traditional relational database is based on relations. When representing temporal attributes, it can generate a large amount of data redundancy. Therefore, it cannot represent uncertain spatiotemporal data in relational database effectively. The uncertain spatiotemporal data model was proposed to improve the ability of data representation, which was consist of classifying different kinds of XML nodes and introducing probability attributes to represent uncertain spatiotemporal XML nodes.

Key words: databases; spatiotemporal data; XML; uncertain; data model

XML 具有可扩展特性, 当对模型增添新的信息时, XML 可以很好地将新数据融入, 使得数据的扩展十分方便. XML 仅仅携带数据, 与内容有关, 是独立于平台和具体应用的, 因此在不同应用程序之间进行数据的交换很方便. 而且 XML 可以提供一种标准化的方法, 种种优点都凸显出了 XML 在互联网迅速发展的今天, 将其运用到数据库建模中将会带来很大的方便^[1-3].

XML 的可扩展性很强, 在 XML 中, 由于标记是可以自定义的, 因此每个人都可以定义属于自己的标记, 使数据可以便利清晰地存储, 进而实现自己的目的. 通过 DTD 文件可以实现文档的共

享, 不同的 XML 文档如果使用的是相同的 DTD 文件, 那么其结构都是相同的, 因而可以使不同组的人无障碍地交流数据. 当收到一份数据时, 可以对比已有的 DTD 文件, 以此来判断收到的数据是否符合规范, 若不符合则将会提示出错. 也可使用 DTD 文件检查自己的数据是否符合定义的规范^[4-5].

XML 已经形成了国际化的标准, XML 的新编码标准可以支持任何编程语言, XML 分析器都可以识别, 因此基于 XML 的各种应用能够处理不同语言的字符. 由此, XML 可以在不同系统之间无障碍地交流数据内容和在整个互联网上自如

地交换数据信息.此外,XML 还可以在不同类型数据、不同系统间交换格式,并进行数据信息的传送,从而使得两个应用进程之间的信息交流、传递信息的工作变得简单多了.XML 的种种特性使得其应用软件的功能十分强大,不仅如此,也使得灵活性很高,这给广大用户提供了更多的方便.

时空数据很复杂,难以直观地描述^[6-7].可以结合 XML 自描述性定义的优点来表述时空数据的信息.进行数据建模时,可以根据时空数据自身的特征来自定义标记,创建模式.进而使其内容独立于具体的应用,这样的做法还有利于数据的重用.编写文档类型定义 DTD,有一个规范化的表达,有了模式对于新定义的时空数据的内容还可以进行有效性验证,确保数据是有效的,避免了后续麻烦.由此,将 XML 的种种特性和不确定时空数据的建模结合起来^[8-10],进而提供更好的服务,是本文要实现的目标.现阶段的研究都是单独对某一方面进行的研究,而将 XML 与不确定时空数据结合进行研究相对来讲还没有进行深入的研究,本文将 XML 与不确定时空数据相结合,借助 XML 可自定义、可扩展、结构化语言的优点,可以更加清楚简单地表述出时空数据,并以概率作为属性,完整地不确定时空数据进行表述.

1 不确定时空 XML 文档节点的分类

XML 文档是有根有序的树形结构.首先对节点进行分类,包括值节点(圆形)、时态节点(正方形)、可能性分布节点(五边形)、根节点(椭圆形)、元素节点(六边形)、不确定性节点(三角形)和属性节点(矩形),如图 1 所示.

1) 根节点(root node):XML 文档树有且仅有一个根节点,根节点是所有元素的父节点,模型中用椭圆进行表示.

2) 元素节点(element node):元素节点可以是其他节点的父节点,而元素节点的父节点只能是元素节点或根节点.普通的元素节点用六边形表示.

3) 属性节点(attribute node):要和它的父元素节点一起使用,在模型中用矩形表示.时间属性在模型中采用正方形表示.

4) 值节点(value node):是叶子节点,可以是数字或是文本,以圆圈来表示.

5) 不确定节点(uncertainty node):对不确定时空数据,将其视为元素,用三角形表示.不确定

时空数据模型中,以概率来表示数据的不确定性,这里定义一个特殊的属性节点(概率属性节点),模型中以五边形表示.概率的取值范围为 $[0,1]$.

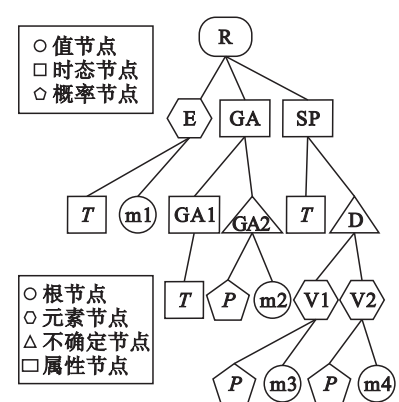


图 1 XML 文档中节点的分类
Fig. 1 Classifications of nodes in XML documents

2 基于 XML 不确定时空数据模型

XML 数据模型基本的结构是树.因此可以先把每个不确定时空数据看成一个空的 XML 文档树,然后依据不确定时空数据的特点,逐步构架出不确定时空数据的表示模型.

定义 1 基于 XML 的不确定时空数据(USP)表示为五元组: $USP = \{OID, ATTR, UP, UM, UT\}$,其中,OID 表示不确定时空数据变化历史,它不仅描述了变化的类型,还说明了对象由什么变化而来,要变化成什么对象;ATTR 用来表示不确定时空数据的属性,将属性分为一般属性(general)和空间属性(spatial),其属性可以是一个或多个,可以是确定的属性也可以是不确定的属性;UP 表示的是不确定时空数据的位置;UM 表达了不确定时空数据的运动趋势,包括方向和数值;UT 表示时间.

2.1 时空对象的变化

由一个时空对象变化成为另一个时空对象,并且要指出变化的类型.

定义 2 时空对象的变化可以用一个四元组表示: $CH(O) = \{pre, suc, type, T\}$,其中,pre 表示之前的时空对象,即由何对象变化而来;suc 表示变化之后的时空对象;type 表示变化的类型:创建(create)、分裂(split)、合并(merge)、消亡(eliminate),分别定义为 c,s,m,e;T 表示时间,可将其视为属性,可以表示时刻也可以是区间,如果是表达区间,则增加两个属性 T_s 和 T_e 作为有效时间的起始时间和结束时间.

给出不确定时空对象变化历史的结构图,如

图 2 所示:

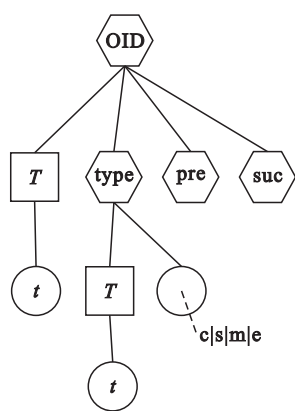


图 2 不确定时空对象的变化历史
Fig. 2 Changing history of uncertain spatiotemporal objects

2.2 不确定时空对象的一般属性

不确定时空对象的属性可以分为一般属性 (general attribute) 和空间属性 (spatial attribute)。这里首先讨论时空对象的一般属性。一般属性也可以分为确定和不确定两种情况。例如某地块的属性有唯一的编号和所有者,编号是 001 保持不变,但是地块的所有者可能会发生变化,比如说所有者可能是 May 的也可能是 Ann 的,这时候就是不确定的。时空对象一般属性的结构图如图 3 所示。

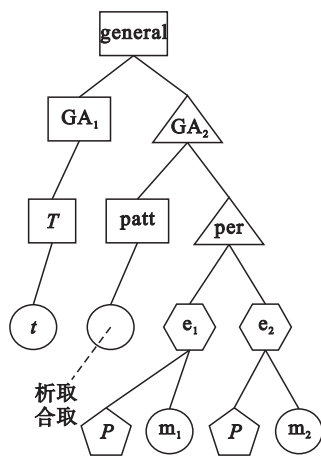


图 3 不确定时空对象的一般属性
Fig. 3 General attributes of uncertain spatiotemporal objects

图 3 中 patt 表示不确定属性的取值类别,分为合取、析取两种,per 表示可以取的元素是 e_1 , e_2 。合取是取二者的并集,而析取是指取二者其中一个。在本例中土地的所有者只能是二者中的一个,因此 patt 类别为析取 (disjunctive)。其 XML 文档表示如下:

```
<land id = "001" Ts = "0" Te = "Now" >
```

```
< owner >  
< dist patt = "disjunctive" >  
  < val pro =0. 8 > May </val >  
  < val pro =0. 7 > Ann </val >  
</dist >  
</owner >  
</land >
```

在上述 XML 文档中,地块的主人可能是 May 或者是 Ann,但只能属于一个人,所以类型选择 disjunctive,并为选择的所有者附加一个概率属性,用以表达作为地块属主的概率是多大。

2.3 不确定时空对象的位置

空间数据结构是对空间数据元素之间结构关系的描述,空间数据模型存储方式分为两大类:矢量数据和栅格数据。本文采用矢量型数据,将“面”实体通过坐标的形式表示。对于面的描述,可以对空间对象的实际形状进行抽象化和近似化,这里采用最小外接矩形 (MBR) 来近似表示一个时空对象。

定义 3 任意时刻 t ,时空对象的位置可以用一个面表示,可用最小外接矩形来近似表示,即 $\{(x_a, y_a), (x_b, y_b)\}$ 。其中, $x_a = \min \{x_1, x_2, \dots, x_n\}$, $x_b = \max \{x_1, x_2, \dots, x_n\}$, $y_a = \min \{y_1, y_2, \dots, y_n\}$, $y_b = \max \{y_1, y_2, \dots, y_n\}$ 。

由于在传输过程中可能会产生诸多不确定因素,进而会导致数据的失真,对于不确定时空数据,将以概率 P 来表述对象出现在某一个位置的的概率,位置的结构图如图 4 所示。

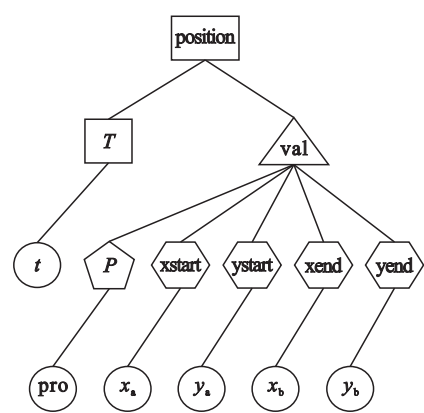


图 4 不确定时空对象的位置
Fig. 4 Position of uncertain spatiotemporal objects

2.4 不确定时空对象的运动

时空对象的运动用来描述运动的方向,以及在各方向上运动的数值大小。考虑时空对象的运动,即该对象分别在 x 轴、 y 轴的运动趋势,以及改变的大小。不确定时空对象运动的结构图如图

5 所示。

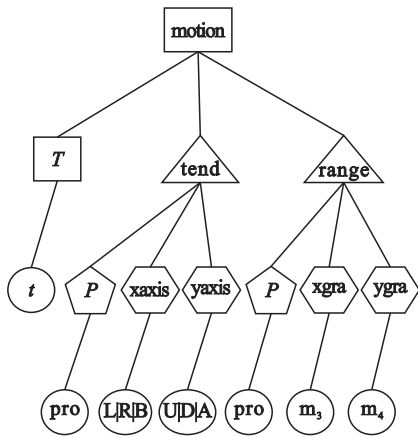


图 5 不确定时空对象的运动

Fig. 5 Motion of uncertain spatiotemporal objects

定义 4 设在 t_1 时刻时空对象的最小外接矩形表示为 $\{(x_a, y_a), (x_b, y_b)\}$ ，在 t_2 时刻时空对象的最小外包矩形表示为 $\{(x_c, y_c), (x_d, y_d)\}$ 。假设时刻 $t_2 > t_1$ 。

定义 5 如果 $1/2(x_a + x_b) - 1/2(x_c + x_d) > 0$ ，则在 x 轴的方向上时空对象是向 x 轴的负方向运动，向左 (L) 运动。如果 $1/2(x_a + x_b) - 1/2(x_c + x_d) < 0$ ，则在 x 轴的方向上时空对象是向 x 轴的正方向运动，向右 (R) 运动。如果 $1/2(x_a + x_b) - 1/2(x_c + x_d) = 0$ ，则在 x 轴的方向上时空对象是没有动 (B) 的。如果 $1/2(y_a + y_b) - 1/2(y_c + y_d) > 0$ ，则在 y 轴的方向上时空对象是向 y 轴的负方向运动，向下 (D) 运动。如果 $1/2(y_a + y_b) - 1/2(y_c + y_d) < 0$ ，则在 y 轴的方向上时空对象是向 y 轴的正方向运动，向上 (U) 运动。如果 $1/2(y_a + y_b) - 1/2(y_c + y_d) = 0$ ，则在 y 轴的方向上时空对象是没有动 (A) 的。引入概率属性表示当不确定时空对象在 y 轴上运动发生变化的概率。

定义 6 分别在 x 轴、 y 轴两个方向上计算空间对象大小的变化，在 x 轴方向上空间对象大小的变化率 $xgra = (x_d - x_c) / (x_b - x_a)$ ，即在 x 轴上扩大或缩小的倍数；在 y 轴方向上空间对象大小的变化率为 $ygra = (y_d - y_c) / (y_b - y_a)$ ，即在 y 轴方向放大缩小的倍数。

对于不确定时空数据加入概率属性，并为该属性赋值，表示出 t_2 时刻，对象的位置在此处的可能性，在定义时可以设置概率属性的默认值为 1，则当概率为 1 时可以省略不写。下面举一个时空对象运动的例子，如图 6 所示。

由图 6 可知： $1/2(1 + 2) - 1/2(3 + 6) = -3 < 0$ ，在 x 轴的运动方向为向右，用 R 表示；

$1/2(2 + 4) - 1/2(1 + 3) = 1 > 0$ ，在 y 轴的运动方向为向下，用 D 表示； $xgra = (6 - 3) / (2 - 1) = 3$ ， x 轴方向的长度变为原来的 3 倍； $ygra = (3 - 1) / (4 - 2) = 1$ ， y 轴方向上的长度变为原来的 1 倍。根据这些值可以找到 t_2 时刻对象的位置。由于传输过程中会存在数据的失真、丢失等问题，加入概率属性表示数据的不确定性。

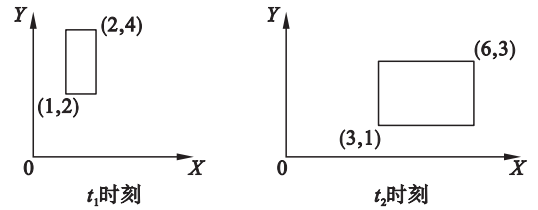


图 6 不确定时空对象运动的例子

Fig. 6 An example of motion of uncertain spatiotemporal objects

基于 XML 将图 6 所举的不确定时空对象运动的例子表示如下：

```
< motion Ts = "t1" Te = "t2" >
  < tend >
    < val pro = 0.8 >
      < xaxis > R < xaxis >
      < yaxis > D < /yaxis >
    < /val >
  < /tend >
  < range >
    < val pro = 0.7 >
      < xgra > 3 < /xgra >
      < ygra > 1 < /ygra >
    < /val >
  < /range >
< /motion >
```

3 基于 XML 不确定时空数据的结构

此前已经详细地介绍了不确定时空对象的变化历史、一般属性和空间属性等并给出了结构图。在此基础上基于 XML 组合成不确定时空数据模型的结构图，如图 7 所示。

从图 7 中可以看出，每一个节点都有自己的时间属性，这样可以更清楚地表达空间对象随时间变化的模型结构，在进行 XML 文档的表述时，将时间作为一个属性，很容易看出来在某时间发生的变化情况，同时也将概率作为一个属性，表述时空数据的不确定性。

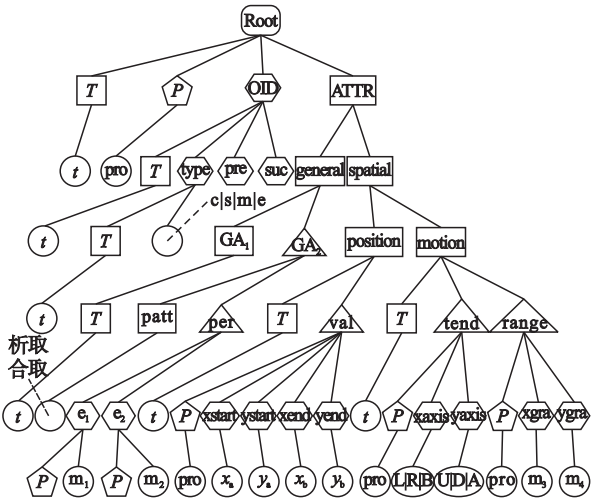


图 7 不确定时空 XML 数据模型的结构
Fig. 7 Structure of uncertain spatiotemporal XML data model

4 结 语

本文介绍了将 XML 的优点与不确定时空数据的表示进行结合. 将整体结构的模型分解成 5 个小部分分别进行详细的分析,并在最终进行整合. 在表示模型中,把时态这个因素作为各个元素的属性融入结构模型中,来体现元素和时间的关系. 时空对象的不确定性种类很多,例如传送过程中可能会有信号的失真、丢失等问题,这就会导致数据的不确定性. 为了表述数据的不确定性,本文加入了一个概率的属性来予以表述. 给不确定的时空数据,加上一个概率属性,这样可以把不确定的数据表示出来,形成一个完整的不确定时空数据的模型.

参考文献:

[1] Jiang X J,Zhang Y P,Pi D C. An XML-based data model for moving object database [J]. *Communications in Computer and Information Science*,2010,118:19 – 28.

[2] Franceschet M, Montanari A, Gubiani D. Modeling and validating spatio-temporal conceptual schemas in XML schema [C]//*Proceedings of the 18th International Conference on DEXA*. Regensburg,2007:25 – 29.

[3] Liu X H,Wan Y C. Storing spatio-temporal data in XML native database [C]//*Proceedings of the 2nd International Workshop on Database Technology and Applications*. Wuhan,2010;1 – 4.

[4] Chen S K,Lo M L, Wu K L, et al. A practical approach to extracting DTD-conforming XML documents from heterogeneous data sources [J]. *Information Sciences*,2006,176(7):820 – 844.

[5] Emadi M, Rahgozar M, Ardalan A, et al. Approaches and schemes for storing DTD-independent XML data in relational databases [J]. *Transactions on Engineering, Computing and Technology*,2006,13:168 – 173.

[6] Pelekis N, Theodoulidis B, Kopanakis I, et al. Literature review of spatio-temporal database models [J]. *The Knowledge Engineering Review*,2004,19(3):235 – 274.

[7] Nandal R. Spatio-temporal database and its models;a review [J]. *Journal of Computer Engineering*,2013,11(2):91 – 100.

[8] Emrich T, Kriegel H P, Mamoulis N, et al. Querying uncertain spatio-temporal data [C]//*Proceedings of 28th ICDE*. Washington D C,2012:354 – 365.

[9] Senellart S, Abiteboul S. On the complexity of managing probabilistic XML data [C]//*Proceedings of the 26th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*. Beijing,2007:283 – 292.

[10] Emrich T, Kriegel H P, Mamoulis N. Indexing uncertain spatio-temporal data [C]//*Proceedings of the 21th CIKM*. Sheraton,2012:395 – 404.

(上接第 1545 页)

[9] Šidlauskas D,Šaltenis S, Jensen C S. Processing of extreme moving-object update and query workloads in main memory [J]. *The Very Large Database Journal*,2014,23(5):817 – 841.

[10] Šidlauskas D,Šaltenis S, Jensen C S. Parallel main-memory indexing for moving-object query and update workloads [C]//*Proceedings of the Management of Data*. Scottsdale:ACM,2012:37 – 48.

[11] Moon B,Jagadish H V, Faloutsos C, et al. Analysis of the clustering properties of the Hilbert space-filling curve [J].

IEEE Transactions on Knowledge and Data Engineering,2001,13(1):124 – 141.

[12] Šidlauskas D, Ross K A, Jensen C S, et al. Thread-level parallel indexing of update intensive moving-object workloads [C] // *Proceedings of Advances in Spatial and Temporal Databases*. Berlin:Springer,2011:186 – 204.

[13] Sarma A D, Gollapudi S, Najork M, et al. A sketch-based distance oracle for web-scale graphs [C] // *Proceedings of ACM International Conference on Web Search and Data Mining*. New York:ACM,2010:401 – 410.