

GIS 时空数据模型研究综述^{*}

Overview of Spatio-Temporal Data Models

舒红¹ 陈军^{1,2} 史文中³⁽¹⁾武汉测绘科技大学测绘遥感信息工程国家重点实验室 武汉 430070)⁽²⁾国家基础地理信息中心 北京100044)⁽³⁾香港理工大学土地测量与地理资讯系 香港九龙)

摘要 In this paper, research status on spatio-temporal data models have been reviewed. Moreover, from three aspects of spatio-temporal semantics modeling, spatio-temporal data logical modeling including strategies of spatio-temporal data storage, and formal definition of spatio-temporal data models including description of spatio-temporal topological relationships, current technology and existed problems are examined.

关键词 Spatio-temporal semantics, Spatio-temporal topological relationships, Temporal relational data models, Spatio-temporal data models

1 研究动态

万事万物都处在一定的时空坐标系里。作为客观现实世界抽象和表示的时空数据模型日渐成为计算机专家、GIS专家及其他相关行业专家关注的热点课题。

为引起我国GIS技术领导部门的重视,便于国内学者了解国际研究现状,我们从知名学者及其工作、专题讨论会、组织机构及商品化DBMS与GIS四个方面介绍时空数据模型方向的国际研究动态。

1.1 知名学者及其研究工作

对数据的时态特征建模的研究主要集中在三个方面:数据库专家的时态关系数据库研究;人工智能专家的时态关系研究和地理信息系统专家的时空数据建模研究。

自1984年以来,应人事或案情档案、病历、金融数据管理和分析的强烈应用需求,数据库专家开始对事务数据的时态特征建模与时态数据管理进行了大量的研究。典型的代表有Richard T. Snodgrass、James Clifford与Abdullah Uz Tansel、Christian S. Jensen、Shashi K. Gadia、Arie Segev、N. L. Sarda、Gad Riav、Vassilis J. Tsotras及我国的唐常杰和张

师超等。主要的研究成果为提出了关系表、元组和属性字段时间标记的时态关系数据模型,提出了有效时间(Valid time、World time)、事务时间(Transaction time、Database time)和双时间Bi-time(有效时间和事务时间)三种不同含义的标记时间,提出了时刻(Instant、Time point)、时间区间(Time Period、Time Interval)和间隔时间区间集合(Time Element)三种类型的时间标记。基于特定的时态关系数据模型,发展了相应的数据库理论基础:时态关系代数与时态关系演算、时态关系关键字定义、时态依赖、范式定义与规范的时态关系分解原则。同时,还初步探讨了时态关系数据库管理系统的实现机制:时间索引、时态数据的存放策略、时态数据库查询语言(如HSQL、TQUEL、表格式的Time-Query-by-Example、TSQL2)。

数据模型通过描述实体与实体间关系来抽象表示客观现实世界。时态关系是实体间的一类重要关系。从自然语言中语态和时态词语(过去、现在、将来;之…前,之…后等)的机器理解与自动控制中动作执行的时间先后约束应用角度,人工智能专家对事件间的时态关系与时态逻辑进行了深入研究。典型的代表有:James F. Allen于1983年基于时间区间

^{*}国家自然科学基金及国家教委博士点基金资助项目。舒红 助理研究员,博士,从事时空数据库系统理论和实现技术研究。陈军 教授,从事地理信息系统理论和应用研究。史文中 副教授,从事地理信息系统理论研究和教学工作。

提出了13种事件间的时态关系, M. Vilain 和 H. Kautz 于1986年基于时间点提出了时间点代数, Ma Jixin, Brain Knight 于1995年基于时间点和时间区间提出了30种事件间的时态关系, 而张师超于1995年基于间隔时间区间集合提出了21种事件间的时态关系。事实上, 如果将 James F. Allen 的时间区间视为简单时间对象, 间隔时间区间集合可看作由若干满足某种约束条件(集合里时间区间有序排列, 且任意两时间区间无重叠部分)的基本时间对象构成的复杂时间对象。基于复杂时间对象的时态关系可化解为基于简单时间对象间的时态关系, 而 M. Vilain, H. Kautz, Ma Jixin 和 Brain Knight 提出的时间对象主要用于微观、高速运动世界里事件间时态关系的刻画。

地理信息系统作为一种采集、存储、管理、分析与显示与地球表面及其附近空间相关数据的计算机系统, 70年代发展基于文件系统的空间数据模型, 80年代发展基于混合系统(管理图形数据的文件系统和属性数据的关系数据库系统)的空间数据模型, 90年代的趋势是发展基于数据库系统(对象-关系数据库系统、面向对象数据库系统)的空间数据模型。地籍变更、海岸线变化、土地城市化、道路改线、环境变化管理等应用领域需要保存并有效地管理历史变化数据, 以方便将来重建历史状态、跟踪变化、预测未来。时空数据模型的研究起步于90年代, 是空间数据模型的一个重要研究分支。工作在该方向的地理信息系统专家主要有: Gail Langran, Donna J. Peuquet, Michael F. Worboys 及陈军等。其中, 从时空数据存储空间管理的角度, Gail Langran 提出了文件系统支撑下的四种时空数据模型: 时空立方体模型、快照序列模型、基态修整模型及时空复合模型。陈军教授提出了非一范式关系数据库方法支撑下的非一范式关系时空数据模型。而 Michael F. Worboys 建议使用面向对象数据库技术支持时空数据建模。从时空语义理解的角度, Donna J. Peuquet 将时间的语义理解为事件序列, 提出了事件驱动的时空数据模型 TRIAD。此外, Max J. Egenhofer 和 Andrew U. Frank 对空间拓扑关系的渐变规律、时空概念理解做了一些探讨。

1.2 专题讨论会

近十年来, 国际上就时空数据模型及其相关技术问题举行了数次专题讨论会, 比较典型有的:

●1993年于美国阿灵顿、得克萨斯; 1995年于瑞士苏黎世由 James Clifford, Richard T. Snodgrass 组

织的“时态数据库”国际讨论会;

●1993年由 Richard T. Snodgrass 组织的“时态数据库基本框架”国际讨论会;

●1990年1月于美国奥罗诺由 Max J. Egenhofer 组织的“GIS 中的时态关系”国际讨论会;

●1992年9月于意大利比萨由 Max J. Egenhofer, Andrew U. Frank 组织的“GIS 中的时空推理”欧美学者国际学术研讨会;

●1993年5月于美国加州 Lake Arrowhead 由 NCGIA 发起的“地理空间中的时间”国际讨论会;

●1995年9月于中国武汉测绘科技大学测绘遥感信息工程国家重点实验室由陈军教授组织的“三维、时态及动态空间数据建模”国际学术研讨会;

●1997年8月于香港理工大学由陈军教授等倡议举行的“动态及多维 GIS”国际讨论会;

此外, 在 VLDB, ISPRS 等国际会议上最近也设有“时态及时空数据库”的讨论专题。

1.3 研究组织与机构

以时态或时空数据建模作为主要研究方向的国际组织机构及领导人物有:

●ISO SQL3制定委员会设立 SQL/MM 第三部分“时空数据类库”工作小组, Richard T. Snodgrass 任 SQL3草案的 SQL/Temporal 部分制定委员会主席暨 TSQL2语言设计委员会主席;

●Richard T. Snodgrass 和 Christian S. Jensen 合作主持 TIMECENTER (“在传统和新型 DBMS 上增加功能支持时态数据库应用”国际研究中心)工作;

●ISO/TC211设立第四工作组 WG2(地理空间数据模型及空间操作算子)时态子模式专题 W1 小组, 美国 Charles Roswell 任小组组长;

●陈军教授于1996年6月当选为 ISPRS IV/3 “空间数据库的时态特征与数据更新”工作小组组长;

●欧洲 DG XII 科学、研究和开发委员会资助的, 作为“欧洲时空数据库系统研究人员培养与交流网络”活动之一的 CHOROSCHRONOS(1996年-1999年), 由雅典国家技术大学的 T. Sellis, M. Stefanakis, Y. Theodoridis 和 M. Vazirgiannis 负责日常协调管理工作, 旨在增强欧洲地区的时空数据库研究力量;

●美国国家地理信息分析中心 Main 大学分部于1994年将“GIS 中的时空推理”列为第十大最有潜力的研究课题, 并由 Max J. Egenhofer 领导该研究

小组。

此外,在 Arizona 大学、丹麦 Alborg 大学、英国 Keele 大学、Main 大学、Pennsylvania 大学、希腊雅典国家技术大学、Maryland 大学、加拿大 Laval 大学、荷兰 ITC、四川联合大学计算机系、武汉测绘科技大学测绘遥感信息工程国家重点实验室等单位也纷纷设立了时态、时空数据库的研究方向和自己的研究小组。

1.4 商品化 DBMS 与 GIS

随着计算机硬件的飞速发展及 DBMS 和 GIS 应用领域的扩大,部分厂家已在其系统里加进了时空数据建模、管理和分析的功能。例如:

- Informix Illustra 研制用于金融领域时态数据建模的“时间序列数据插片”,提供时间序列 Time series 和日历 Calendar 两个数据类型与10个特定的操作函数。同时,还研制“二维和三维空间数据插片”(1994年),Mapinfo 联合研制“地理编码(Geocoding)数据插片”;

- Oracle 的空间数据选项模块 SDO;ESRI 的统一管理大容量的图形和属性数据的空间数据库引擎 SDE;

- CA-OpenIngress 的对象管理扩充器 OME 和空间对象库 SOL;SIRO-DBMS 的空间数据管理工具箱(1989年);

- SYBASE 与 Vision 国际公司联合推出的空间查询服务器;

- NASA 采用 Informix 的 Universal Server 构造了一个四维系统,处理每天24小时来自5颗卫星的约1TGB的数据,系统运行良好;

- 国内外 GIS 厂商,基于 Microsoft 的 ODBC API 上开发的 Map Object Server, Spatial Object Server。

此外,Oracle、Informix、Sybase、Ingres、Microsoft SQL Server、IMB DB2等纷纷提供 Date、Time、Datetime、Timestamp、BLOB、Varchar、Binary、Image 等基本数据类型及相应的操作,提供用户定义数据类型、用户定义函数手段与数据插片技术。这些预示着时态、时空数据建模的支持会成为下一个时段信息系统软件竞争市场的一个重要指标。

2 现有成果及问题

通用数据库管理系统里采用的层次、关系或对象数据模型是较广范围的客观世界里事物(最初主要是企业事务)的普遍特征与规律的一种模拟表示。

而时空数据模型则是客观地理世界具体范围内事物的基本特征与规律的一种模拟表示。有时,人们将通用 DBMS 里的数据模型称为通用数据模型,而将时空数据模型称为 GIS 应用数据模型。

回顾过去近十年的时空数据建模工作,在时空语义建模、逻辑建模以及模型的形式化描述方面取得的成果还相当有限,存在许多不足或需要改进的地方。

2.1 时空语义

语义是数据的一种含义,是人类对客观世界本质特征与规律正确认识和理解后形成的一种概念。如何将这种概念借助于某种手段(如关系表、对象聚集和概括等)直观地放入数据库即是语义建模的根本任务。如果说数据库里的信息包括直观信息和隐含信息两部分,则直观信息即为数据模型的语义,而隐含信息为应用领域语义。自然地,模型的语义越强,数据库的信息会越丰富。

较之层次、关系或对象等结构化数据模型,时间和空间是一种更具体的概念,时空数据建模的根本目的在于时间和空间语义概念的建模。然而,过去的工作多集中在数据的存储管理与操作等计算机实现方面。这主要由于时空数据模型的研究受数据库专家思想的影响。毕竟,对时态特征建模研究,数据库专家起步早、研究较深入。近年来,Max J. Egenhofer、Andrew U. Frank (1992)、Donna J. Peuquet (1993)、Christophe Claramunt (1996)、Marius Thériault (1995)和史文中 (1997)等 GIS 学者开始关注时空语义建模,提出了事件驱动和面向过程的时空数据模型。但总的看来,对究竟什么是地理世界中的物理时间和空间?什么是人们地学认知世界中的时间和空间概念?什么是地理信息系统模拟的时间和空间概念?时间和空间概念如何结合及其结合的内在含义是什么?时空数据模型应该表现及如何表现哪些实用的时空概念?过去有关时空数据语义建模理论研究存在不足。

来自哲学和认知学的两种时间观点:事件序列的时间观与数量尺度的时间观。根据这两种观点可发展两种 GIS 中的时态数据模型:事件驱动的时态数据模型与时间标记的时态数据模型(舒红,1998)。空间和时间是运动物质存在的两种基本形式。在地理信息系统中,空间刻画了地理实体的空间位置、空间分布与空间相关性;时间刻画了地理实体的存在时间、变化状况、时间相关性。空间、时间和属性是地理实体的三个基本特征。脱离地理实体这个中心,单

独谈论空间、时间、时间和空间的关系已超出了信息系统设计的思考范畴。从信息系统的观点出发,事件通过相关状态的差异(变化)来表现,其含义由用户结合具体应用领域解释获得。总结各种不同含义的地理事件或地理变化类型是一项很有意义的工作,是基于事件序列的时间概念的地理对象模拟与推理的基础。其意义如同认识地理实体的空间分布特性和地球表面的空间组织规律有助于空间数据模型的建立一样。信息系统里的时间尺度一般是有限的、有向的及离散的。时空数据模型里的语义包括时间语义、空间语义和属性语义。地理实体的状态由地理空间状态和地理属性状态构成,则地理变化可划分为地理实体的空间状态变化、属性状态变化和空间属性状态变化。深层的时间语义(事件序列)本质上位于空间语义和属性语义之上(图1)。



图1 时空数据模型的语义层次

明确了时间语义,我们也就明确了建立时空数据模型的基本作用:帮助人们理解地理实体的变化过程,确定地理实体状态的时间标记,从而避免对时空数据模型的盲目研究。

2.2 逻辑建模

时空数据逻辑建模的内容包括:时空数据类型及其数据结构实现、时空数据操作及时空完整性约束三部分。一个有效的时空数据模型必须具有如下特点:

1)降低冗余存储。如减少空间公共边、时间不变数据的重复存储。

2)支持各种复杂时空对象的构造。时空对象通常具有层次性(聚集层次和概括层次)、变长字节存储性、时空有序性等结构特点。模型必须提供有效的手段支持如此结构特点的对象数据建模。

3)表达丰富的时空语义。尤其表达空间结构和空间拓扑关系等空间语义,表达事件序列(变化过程)和时间相关性等时间语义。

4)兼顾检索效率和用户的应用要求。需要折中考虑数据的空间内聚性和时间内聚性,合理选择时间标记对象的粒度。如全局状态的快速提取、局部变

化的快速发现、频繁使用和较少使用的数据的合理放置。一般地,当前数据被放置于本地高速硬盘上,历史数据被放置于远程磁带或光盘上,并分别提供索引。

Gail Langran 建议的四种时空数据模型中,时空立方体模型将有效时刻标记在空间坐标点上,快照序列模型将有效时刻标记在全局空间状态上,基态修整模型将有效时间标记在两状态的变化差值上,而时空复合模型将有效时间标记在人为投影在某一时刻形成的复合图形单元上。一般情况下,时空立方体模型中存在时间标记本身的冗余存储,而快照序列模型中存在不变空间状态数据的大量冗余存储。基态修整模型和时空复合模型较为节省存储空间,但差值状态数据的获取与复合图形单元的形成困难,且带来了全局状态恢复时需进行大量零碎状态重组,破坏了地理实体的时空结构完整性。

非一范式关系时空数据模型仅注意到地理实体的时变序列数据结构特点,对时态属性数据建模较为有效,但它忽略了空间数据的层次性、变长字节存储性和空间有序性特点,支持时态空间数据建模能力有限。Donna J. Peuquet 的事件驱动的时空数据模型显式存储事件(状态变化)序列,模型顾及了时间语义的表现,但类似于基态修整模型,管理检索变化数据很困难。

值得一提的是,Michael F. Worboys (1994)、Donna J. Peuquet (1995)、舒红(1997)等首先关注面向对象技术在时空数据建模中的运用。尽管面向对象技术在建模概念、理论基础和实现技术上还没有达成共识、不够成熟。但支持复杂结构对象建模是任何对象数据模型所必须具备的基本能力。这一点,勿容置疑。时空数据是一类复杂的数据;使用面向对象数据技术建模是很自然的。1997年香港 ISPRS 国际讨论会上,作者首次提出时态对象代数的概念,并建议将时空数据建模放在三个不同层次上:复杂结构对象的基本时态特征建模、通用时空对象建模、应用领域时空数据建模。从数量尺度的概念理解时间,时间是一维、有向的,较三维空间数据结构简单,故将对象时态特征在系统底层模拟是合理的。时态对象代数与 Richard T. Snodgrass、James Clifford、Shashi K. Gadia、Christian S. Jensen 和张师超等人发展时态关系代数的思路是一致的。时态关系代数的基本思路是通过在关系结构上显式化时变语义来进行简单结构时态数据建模,而时态对象代数的基本思路则是在对象结构上显式化时变语义来进行复杂结构

时态数据建模。

总结一套兼顾时空语义表现、时空存储存取效率、不同应用需求的时空数据逻辑建模指导原则,发展规范化的面向对象时空数据模型是必要且紧迫的。

2.3 模型的形式化定义

模型的形式化定义(数学的或逻辑的定义)是数据模型研究的另一个重要方面,是数据库理论的一个重要组成部分。模型的形式化定义主要包括对象结构的形式化定义和对象基本操作的形式化定义,时空数据模型的形式化定义还包括时空关系的形式化定义。

基于元组时间标记和基于属性字段时间标记的时态关系代数与时态关系演算已有较完善的结论,并且基于各种形式化模型,发展了相应的时态数据库查询语言。着眼于复杂数据结构建模的对象代数、着眼于空间几何构成建模的流形拓扑和超图等理论正被国内外学者努力用于不同抽象层次的时空数据模型形式化定义。

时空关系的形式化定义必须慎重考虑三个基本问题:1)语义的合理性问题,时空关系概念必须符合人类对客观世界的认识并有助于人类改造客观世界的需要;2)数学形式的完备性和唯一性问题;3)存储表达问题。

James F. Allen 给出的基于时间区间的13种时态关系成为时态关系,甚至空间关系形式化描述的里程碑。Max J. Egenhofer 基于点集拓扑理论的空间拓扑关系形式化定义得到普遍的认可,ISO SQL/MM 工作草稿里将其列为空间关系描述的理论依据是一有力证明。在空间拓扑关系描述的四元组4I框架基础上,扩展的九元组9I(对象内部、边界和外部交矩阵)、空间对象维数值、对象维数和背景空间维数差值等更细微区分空间拓扑关系的理论框架已建立。进一步,Eliseo Clementini 等还给出了便于用户理解且等价于点集拓扑空间拓扑关系集合的空间拓扑关系最小逻辑谓词描述集。虽然,时态关系和空间拓扑关系的形式化定义分别有了一定的结论,但大家广泛关注的时空拓扑关系却还未有一个明确的定义。时空拓扑关系的作用和真实含义是什么?如何用数学工具完整地描述它?怎样在数据库里有效存储它?这些问题都亟待解决。

结束语 本文仅就时空语义建模、时空数据逻辑建模及时空数据形式化模型三个方面综述了已有的技术成果及存在的问题。关于时空数据物理建模

(包括系统结构、物理数据结构、数据索引等)方面的技术与问题,拟另文介绍。

最后,作者对丹麦 Alborg 大学计算机系从事时态数据库研究的 Christian S. Jensen 教授和在荷兰国际航天测量与地球科学学院 ITC 从事时态地理信息系统研究的程涛博士表示感谢,感谢他们对作者研究工作的鼓励和支持。

参考文献

- [1] Gultekin Ozsoyoglu et al., Temporal and Real-Time Databases: A Survey, IEEE Trans. on Knowledge and Data Engineering, 7(4)1995
- [2] James F. Allen, Maintaining Knowledge about Temporal Intervals, CACM, 26(11)1983
- [3] 张师超,时态关系代数与元组演算的等价性,计算机学报,16(12)1993
- [4] 张师超,基于间断区间的时态知识表示,软件学报,5(6)1994
- [5] Gail Langran, Time in Geographic Information Systems, Taylor & Francis Ltd., 1992
- [6] M. F. Worboys, Object-Oriented Approaches to Geo-Referenced Information, Int. J. Geographical Information Systems, 8(4)1994
- [7] Donna J. Peuquet et al., An Event-Based Spatio-Temporal Data Model (ESTDM) for Temporal Analysis of Geographical Data, Int. J. Geographical Information Systems, 9(1)1995
- [8] 陈军,陈尚超,唐治峰,用非第一范式关系表达 GIS 时态属性数据,武汉测绘科技大学学报,20(1)1995
- [9] 舒红,陈军,杜道生,周勇前,面向对象的时空数据模型,武汉测绘科技大学学报,22(3)1997
- [10] H. Shu et al., A Temporal Object Algebra to Support Spatio-Temporal Data Modeling, In Proc. of Intl. Workshop on Dynamic and Multi-dimensional GIS, Hongkong, 1997. 8
- [11] Matthew Y. C. Pang et al., Development of Process-Based Data Model for Temporal Geographical Information Systems, Proc. of GIS AM/FM ASIA'97 & Geoinformatics'97, Taipei, Taiwan, May 26-29, 1997
- [12] Max J. Egenhofer et al., Point-Set Topological Spatial Relations, Int. J. Geographical Information Systems, 5(2)1991
- [13] Eliseo Clementini, et al., A Small Set of Formal Topological Relationships Suitable for End-User Interaction, Advances in Spatial Databases, Springer-Verlag, 1991
- [14] Ma Juxin et al., A General Temporal Theory, The Computer Journal, 37(2)1994