

车辆监控数据仓库中时空立方体 数据模型的设计与实现

胡继华¹ 程智锋¹ 詹承志¹ 唐 巍¹

(1 中山大学工学院智能交通中心,广州市大学城外环东路 132 号,510275)

(2 广东省智能交通系统重点实验室,广州市市新港西路 135 号,510006)

摘要:分析了车辆监控数据和应用的特点,设计了车辆监控的时空索引机制和车辆监控数据仓库。利用广东省重点车辆监控数据实现了车辆监控数据仓库,并与传统数据库管理方式的效率进行对比分析表明,设计的时空数据仓库能够满足海量车辆监控数据管理和分析的要求。

关键词:时空立方体;数据仓库;车辆监控;双向链表

中图分类号:P208

车辆监控是对车辆的运营状态和违法行为进行监控,接收车辆报警信息,并通知交通管理部门出警或应急救援。车辆监控平台定时采集大批量的车辆位置状态数据并进行分析,所采集数据量非常庞大。同时,车辆监控和时间关系密切,是典型的时空过程。传统的数据库难以对车辆监控数据进行有效的管理和分析。

时空数据模型是管理时空数据的有效方法。时态数据库^[1]、面向对象数据库技术^[2]常应用在时空数据模型的构建上。在大量时空数据索引方面,常用的有基于事件的时空处理模型和 R 树索引方法^[3-4]。Vania 等^[5]则针对移动对象轨迹检索问题,设计了一种时空数据挖掘的查询语言。Wonik 等^[6-7]首次提出了采用时空立方体模型对移动对象建立索引,证明了时空立方体处理方式比 R 树具有更高的处理效率。薛存金等^[8]提出了面向过程的时空数据模型,并以海洋时空过程为研究对象构建了原型系统平台。孟令奎等^[9]讨论了一种基于地理事件时变序列的时空数据模型,并在 GIS 平台上进行了拓展。俞肇元等^[10]根据 Clifford 代数空间构建思想扩展了现有时空数据模型,实现了时间、空间和属性的一体化表达。李清泉等^[11]讨论了时空一体化 GIS-T 数据模型的基本框架。王卫京等^[12]在时空立方体模型基础上提出了通过时空立方体建立轨迹版本的思

想。考虑到车辆监控的空间范围大,时间连续性要求高等特点,本文采用时空立方体单元所在的空间网格经纬度及时间作为其 ID,建立空间网格+时间+双向链表的时空索引机制,然后使用广东省重点车辆监控数据,设计实现了时空数据仓库。

1 时空立方体模型

1.1 时空立方体单元

时空立方体单元是以规则的空间网格为底,相应的时间段为高组成的方形柱体,是数据存储的基本单元。在时间维度,则按照时间段进行分区,记为 ΔT 。 $\Delta S \times \Delta T$ 代表了一个三维空间,即一个小的时空立方体单元,记为 ΔC 。在空间平面中,每个小的网格单元都具有一个固定的网格编码值 g ,用于对数据按照空间范围进行索引,其定义如下:

$$g = f(x, y, r) \quad (1)$$

在时间维度单元 ΔT 中,每一个时段 ΔT 都包含多个时间快照。记一个不可再分的时间片单元为 $E[g, t]$,以 ΔC 为时空间隔的时空立方体单元为 $C[g, b]$,一个时间维度单元间隔 ΔT 构成的空间平面数据集为 $B[b]$ 。其中, t 和 b 分别为时间片和时空立方体单元的时间尺度大小。时空立方体 ST-Cube(spatiotemporal-cube) G 就可以被定义为

收稿日期:2012-11-08。

项目来源:国家自然科学基金资助项目(40971098);广东省 2011 年度安全生产专项资金资助项目(2011-118)。

一系列空间平面集合 $B[b]$ 构成的数据集:

$$G = \cup \{ \cup \{ \cup E[g, t] \} \} \quad (2)$$

1.2 时空立方体自适应裂变

时空立方体单元在存储上可对应一个物理表或逻辑表。设每个表的数据存储量的上限为 V , 超过 V 后, 该表数据的管理效率就会急剧降低。

监控车辆对象在不同时刻的区域分布密度有所差异。按照数据量增长速度的快慢, 时空立方体单元的裂变方式可分为 2 分法、4 分法和 16 分法。设车辆详细信息表为数据桶, 记为 Bucket。分裂过程中一个立方体单元按照进一步的空间划分, 变为子立方体单元, 如图 1 中 T_2 时刻 6 号立方体单元发生的变化。此时, 6 号立方体的指针不再是指向具体的属于相应空间范围内的数据, 而是指向进一步分裂得到的小立方体。

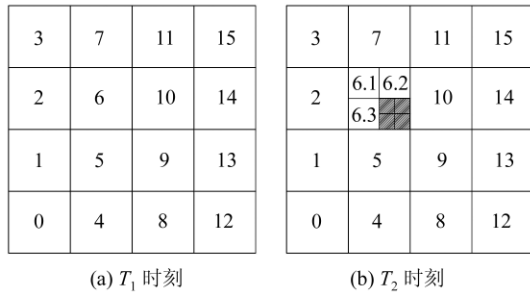


图 1 立方体分裂示意图

Fig. 1 Example of Adaptive Cell

2 数据库设计

2.1 数据入库

1) 历史数据。首先, 按照空间范围划分, 根据数据的经纬度查找相应的网格编号, 将数据录入到相应的数据表当中, 表的命名采用形如“tb_+网格行号+c+网格列号+t+网格层次”的方式, 并将各个表中的相应信息记录到索引表 tb_index 中。其次, 判定每张数据表格是否超过了记录数量阈值, 如果未超过阈值, 可继续填入数据; 若超过阈值, 则采用四分法进行空间网格裂变, 同时在索引表 tb_index 中记录新生成的表的信息, 包括表名、网格编号、起止时间、网格层次等基本信息。由于裂变本质上是少量地增加了时空立方体单元, 同时在另一张表中添加相关立方体记录, 因此总体性能影响不大。最后, 按照车辆车牌进行分车轨迹管理, 即通过对车辆网格变化的判断, 记录车辆进出相应网格的时刻、位置等信息, 建立对车辆轨迹索引的双向链表 tb_link (表 1)。

2) 当前数据。数据接收平台按照设定好的时间间隔将获得的数据存储到当前车辆位置状态的数据表中, 便于对最新车辆状态信息的掌握。同时将信息按照对历史数据的处理方式入库, 主要包括以下 3 个方面的内容: ① 加入到对应网格划分得到的数据表中, 即写入数据到对应的数据区块; ② 判定该数据表中的记录数量是否超过阈值; ③ 根据车辆位置变化, 在需要的情况下更新车辆轨迹的双向链表(图 2)。

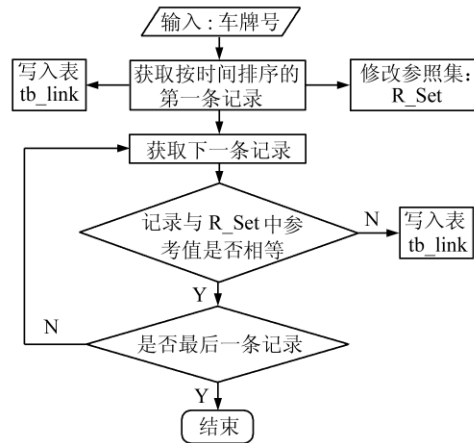


图 2 建立双向链表流程

Fig. 2 Establish the Doubly Link List

2.2 时空索引

按照模型结构设计的数据表主要分为以下 4 个部分: ① 详细历史数据信息表(按设定的粒度建立完成); ② 当前最新车辆信息表; ③ 时空立方体单元信息表; ④ 记录车辆轨迹变化情况的双向链表。其中, 起索引作用的是时空立方体单元信息表和车辆轨迹双向链表。时空立方体单元信息表(tb_index)主要描述了各时空单元表内记录的空间范围、时间范围, 记录数量等总体信息(表 1)。车辆轨迹双向链表(tb_link)主要描述了车辆在各立方体单元中的变化过程, 通过车牌和时间范围确定其对应时间段历史数据存放的数据表(表 2)。

表 1 信息索引表结构

Tab. 1 Structure of tb_index

字段	含义
TABLERNAME	立方体单元表格名称
STARTTIME	所有记录的起始时间
ENDTIME	所有记录的终止时刻
MAPEXTENT	立方体单元跨越的空间范围
GRIDTIER	立方体单元所处的层级
RECORDNUM	立方体单元内所含的记录数量
GRIDROW	立方体单元在空间上网格行号
GRIDCOLUMN	立方体单元在空间上网格列号

表2 车辆行驶轨迹的双向链表

Tab. 2 Doubly Link List Table of Vehicle Trajectory

字段	含义
CARPLATE	记录车辆车牌号码
CHANGRTIME	发生网格跳变的信号传输时间
CHANGETYPE	跳变类型, 0: 进入, 1: 跳出
GRIDROW	发生转变过程的网格行号
GRIDCOLUMN	发生转变过程的网格列号

3 时空数据仓库基本功能

车辆监控时空数据仓库基本功能包括车辆的位置查询、轨迹查询和辖区或道路车辆查询统计。

1) 车辆位置查询和轨迹查询

位置查询分为当前时刻和过去某一时刻两种情况。比较查询时刻与当前时间,如果在更新周期范围内,则直接到当前车辆信息表中找到对应车辆信息,否则,通过双向链表找到其出入网格,在相应信息存储表进行查询。轨迹查询操作首先查询车辆轨迹双向链表,得到查询时段内车辆跨越的时空范围,然后在车辆网格编号对应的表内查找属于查询时段的车辆轨迹信息,最后按照时间顺序综合各表中查到的信息,得到车辆的轨迹。

2) 辖区或道路车辆查询统计

查询某个区域或某条道路上营运车辆分布情况。首先找到与查询时空范围存在交集的立方体单元编号,根据编号从立方体单元结果集中获取详细的数据。并判断时空立方体单元是否存在裂变的情况,如果存在,则通过获得立方体单元的下一级子单元的指针指向找到对应数据。

3) 时空立方体模型查询复杂度分析

传统数据存储方式将所有历史数据存储于同一张数据表中,设该表中共有 N 条记录,则搜索其中某一条记录的检索次数复杂度近似为:

$$O(N) = (1 + N)/2 \quad (3)$$

在建立时空立方体模型后,设划分的立方体单元数为 N_C ,单元内含有记录的最大阈值为 V ,生成的车辆双向链表记录数为 N_L ,则按该模型搜索其中某一条记录的检索次数复杂度近似为:

$$O_{ST-Cube}(N) = (1 + N_L) + \frac{1 + N_C}{2} + V \quad (4)$$

4 实验和验证

4.1 数据和建库

在建立时空数据仓库之前,车辆监控系统常用的数据存储方式是将现有数据与历史数据分开,历史数据为关系数据库分区表。比如按日分

区,分区命名方式为“GPSDATA_yyyymmdd”,其中(yyyymmdd)表示当日日期。分区表间并没有车辆的时空关联信息。从数据中心选取了2010-09广东省重点监控车辆监控数据,总记录数量为16 289 549条。在时空立方体时空单元划分方面采用如下策略:在空间维度上,采用 $\Delta X = 0.3^\circ$, $\Delta Y = 0.3^\circ$ 的单位进行划分;在时间维度上,采用 $\Delta T = 1$ 个月的粒度。此种划分考虑了数据粒度大小。经过测试,100 000条车辆GPS数据点记录是一个可满足查询效率的阈值。建立时空立方体模型,共生成了467个第一层立方体单元,4个第二层立方体单元。每个单元中含有所有1个月时间范围内的车辆运行状态基础数据,同时双向链表中生成了162 700条记录。

最后完成了由车辆轨迹索引到对应数据存储单元索引。根据式(3)、式(4),采用传统数据的存储检索方式和时空立方体模型单元划分方式的平均检索次数分别为8 144 775和262 937。在建立了时空立方体索引之后,记录检索次数复杂度降低,在时空数据查询中更具优势。

4.2 车辆位置查询和历史轨迹回溯

1) 车辆的定位查询

随机选取了10个不同营运车辆,比较时空立方体模型下和常规数据存储方式下单辆车查询处理时间,如图3所示。随机选取6辆营运车辆作为多车同时定位查询实验对象,建立的时空立方体模型前后的平均响应时间分别为0.330 6 s和0.121 8 s。可以看出,建立时空立方体模型下的查询方式在时间花费上有明显的优势。

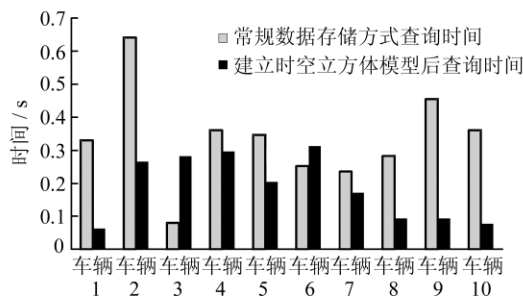


图3 单辆车定位查询时间花费统计图

Fig. 3 Time Spend of Positioning Inquiry

2) 车辆的历史轨迹回溯

实验随机选取10辆营运车辆作为测试对象,时间段则分别选取了1~3 h。建立时空立方体模型的前后单辆车轨迹查询处理时间花费统计如图4。随机选取了6辆营运车作为多车轨迹查询实验对象,同时查询其位于2010-09-19T11:00至2010-09-19T13:00的车辆运行轨迹数据,平均

响应时间分别为 17.493 s 和 10.249 s。通过比较可以看出,建立了时空立方体模型后,轨迹查询时间明显缩短,达到了改进轨迹查询效率的目的。

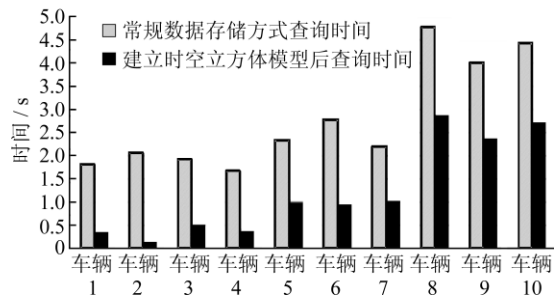


图 4 多辆车轨迹查询时间花费统计图

Fig. 4 Time Spend of Path Backtracking

因此,对营运车辆海量数据建立时空立方体数据模型后,数据查询效率得到明显提高。模型在处理海量营运车辆时态数据方面具有实用性。

参 考 文 献

- [1] Laurini R. Real Time Spatio-Temporal Databases [J]. Transactions in GIS, 2001, 5(2): 81-97
- [2] Worbroys M. Computation with Imprecise Geospatialdata [J]. Computers, Environment and Urban System, 1998, 22(2):85-106
- [3] Wachowicz M. Object-Oriented Design for Temporal GIS[M]. UK:Taylor & Francis, 1999
- [4] Sengupta R, Chen Yan. A Hybrid Spatio-Temporal Data Model and Structure (HST-DMS) for Efficient Storage and Retrieval of Land Use Information[J].

- Transactions in GIS, 2004, 8(3): 351-366
- [5] Bogorny V, Kuijpers B, Alvares L O. A Spatio-temporal Data Mining Query Language for Moving Object Trajectories[R]. Technical Report TR-357, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, 2008
- [6] Choia W, Moon B. Adaptive Cell-based Index for Moving Objects[J]. Data & Knowledge Engineering, 2004, 48:75-01
- [7] Choi W, Kwon D, Lee Sangjun. Spatio-temporal Data Warehouses Using an Adaptive Cell-based Approach[J]. Data & Knowledge Engineering, 2006, 59:189-07
- [8] 薛存金,周成虎,苏奋振,等.面向过程的时空数据模型研究[J].测绘学报,2010,39(1):95-101
- [9] 孟令奎,赵春宇,林志勇,等.基于地理事件时变序列的时空数据模型研究与实现[J].武汉大学学报·信息科学版,2003,28(2):202-207
- [10] 俞肇元,袁林旺,罗文,等. GIS 时空分析系统的 Clifford 代数设计与实现[J].武汉大学学报·信息科学版,2011,36(12):1 397-1 401
- [11] 李清泉,杨必胜,郑年波.时空一体化 GIS-T 数据模型与应用方法[J].武汉大学学报·信息科学版,2007,32(11):1 034-1 041
- [12] 王卫京,翁敬农,樊珂.车辆监控系统中时空数据模型设计与实现[J].计算机工程与设计,2006,27(6): 1 042-1 051

第一作者简介:胡继华,博士,讲师,研究方向为时态 GIS。
E-mail:hujihua@mail.sysu.edu.cn

A Grid-Time-indexed Cube Model for Spatiotemporal Data Warehouse of Vehicle Supervision

HU Jihua¹ CHENG Zhifeng¹ ZHAN Chengzhi¹ TANG Wei¹

(1) Research Centre of Intelligent Transportation System, School of Engineering, Sun Yet-sen University, 135 East Waihuan Road, Guangzhou 510275, China)

(2) Guangdong Provincial Key Laboratory of Intelligent Transportation System, 135 West Xingang Road, Guangzhou 510006, China)

Abstract: Immense amounts of spatiotemporal data generated by vehicle supervision can't be efficiently managed by ordinary databases, due mainly to long query response. This paper proposes a new approach called grid-time-indexed cube, which is a spatial grid indexed, adaptive cube-based and trajectory supported for spatiotemporal data warehouses. Through the extensive performance studies based on spatiotemporal data from main vehicle supervision system of Guangdong province, we observed that grid-time-indexed cube achieved higher query performance than ordinary data storage technologies, under various operational conditions, and it was easily applicable in practice and compatible with ordinary databases.

Key words: spatiotemporal cube; data warehouse; vehicle supervision; doubly linked list

About the first author: HU Jihua, lecturer, Ph.D. majors in temporal GIS and its application in transportation.
E-mail: hujihua@mail.sysu.edu.cn