Journal of Computer Applications
计算机应用
ISSN 1001-9081
CODEN JYIIDU
2018-07-24
http://www.joca.cn

文章编号:001-9081(****)**-0000-00
doi:10.11772/j.issn.1001-9081.2018040749

基于 Hilbert-R 树分级索引的时空查询算法

侯海耀[1]，钱育蓉[1*]，英昌甜[1,2]，张晗[1]，卢学远[1]，赵燚[3]

(1.新疆大学 软件学院，新疆 乌鲁木齐 830008；2.新疆大学 电气工程学科博士后流动站，新疆 乌鲁木齐 830046；
3.新疆大学 信息科学与工程学院，新疆 乌鲁木齐 830046)

(*通信作者电子邮箱 qyr@xju.edu.cn)

摘　要: 针对树形空间索引中多路查询及未考虑时间维索引的问题，提出一种结合时间和聚类结果的 Hilbert-R 树索引构建策略。首先，按照数据采集的周期划分时空数据集，并在此基础上建立时间索引，通过 Hilbert 曲线对空间数据进行分割编码，将空间坐标映射到一维区间；其次，依据数据要素在空间中的分布，采用动态确定 *K* 值的聚类算法，结合聚类结果构建高效的 Hilbert-R 树空间索引；最后，基于 Redis 几种常见的键值数据结构，对时空数据的时间属性和聚类结果构建分级索引。在时空范围及目标矢量对象查询的实验中，与缓存敏感 R+树(Cache Conscious R+Tree,CCR+)相比，所提算法可有效降低时间开销，查询时间平均缩短约 25%，对不同密集型数据具有良好的适应性，可更好地支持 Redis 应用于海量时空数据查询。

关键词: 时空数据；Redis 数据库；聚类算法；Hilbert-R 树；分级索引；时空范围查询

中图分类号:TP311　　**文献标志码：A**

Research on spatio-temporal query algorithm based on Hilbert-R tree hierarchical index

HOU Haiyao[1], QIAN Yurong[1*], YING Changtian[1,2],ZHANG Han[1],LU Xueyuan[1], ZHAO Yi[3]

(1.College of School of Software, Xinjiang University, Urumqi Xinjiang 830008, China;
2.Postdoctoral Research Station of Electrical Engineering, Xinjiang University, Urumqi Xinjiang 830046,China;
3.School of Information Science and Engineering, Xinjiang University, Urumqi Xinjiang 830046,China)

Abstract: Aiming at the problem of multiple query in the tree-space index and not considering the time dimension index, A Hilbert-R tree index construction scheme combining time and clustering results is proposed. First, according to the periodicity of data collection, the spatial-temporal dataset is divided, and on this basis, a time index is established. The spatial data is partitioned and encoded by the Hilbert curve, and the spatial coordinates map to one-dimensional intervals. Secondly, according to the distribution of the feature object in space, proposed a clustering algorithm for dynamically determining *K* value, combined with clustering results to build an efficient Hilbert-R tree spatial index; Finally, based on the several common key-value data structures of Redis, built the hierarchical indexing mechanism of the time attributes-clustering results. Compared with the Cache Conscious R+ tree, the proposed algorithm can effectively reduce the time overhead, and the query time is shortened by about 25% on average in the experiment of space-time and target vector object query. It has good adaptability to different intensive data and can better support Redis to apply to the query of massive spatio-temporal data.

Keywords: spatio-temporal data; Redis database; clustering algorithm; Hilbert-R tree; hierarchical index; spatio-temporal range query

0　引言

矢量时空数据高效组织管理是空间数据应用的关键技术，空间索引是实现矢量时空数据高效检索的关键[1-3]。空间、

收稿日期: 2018-04-12;　修回日期:2018-07-08;　录用日期: 2018-07-09。
基金项目: 国家自然科学基金资助项目(61562086，61462079)；新疆维吾尔自治区教育厅项目（XJEDU2016S035）；新疆大学博士科研启动基金项目（BS150257）；新疆维吾尔自治区教育厅创新团队(XJEDU2017T002)

作者简介: 侯海耀(1990-)，男，山西汾阳人，硕士研究生，主要研究方向：时空数据库索引、遥感图像处理；通讯作者：钱育蓉（1980-），女，山东武城人，博士，教授，CCF 高级会员(23806S)，主要研究方向：网络计算和遥感图像处理；英昌甜(1989-)，女，新疆乌鲁木齐人，博士，主要研究方向：图像处理、 内存计算；张晗(1987-)，男，辽宁本溪人，硕士研究生，主要研究方向：图像处理与模式识别；卢学远(1992-)，男，浙江温州人，硕士研究生，CCF 会员，主要研究方向：图像处理、机器学习；赵燚（1993-），女，新疆克拉玛依人，硕士研究生，主要研究方向：时空数据库索引、遥感图像处理。

时间、属性作为时空大数据的 3 个基本特征[4]，如何描述和表达空间实体及其相互关系的时空变化，成为亟待解决的热点问题。

近年来，树形空间索引技术结合聚类技术来探讨空间索引技术的研究成为热门，对空间数据集进行有效聚类，其结果对构造高效的空间索引具有很大帮助[5]，但从时空查询角度出发，同时考虑所查询数据的时间维特征及各空间要素合理组织的研究较少，还不能很好地满足时空应用中对时空高效查询的需求。文献[6-8]侧重从空间数据分布的特点出发，采用聚类算法与 R 树结合构建索引，提高了空间查找效率；文献[9]提出基于 Redis 的缓存敏感 R+树(Cache Conscious R+Tree，CCR+)，其查询效率较 Oracle 10g 的管理系统有较大的提升，但这些研究均未考虑对时间维建立索引。

基于以上问题，本文利用 Hilbert 曲线较为稳定的离散度分布特点[10]，分割并编码空间要素；利用动态确定 *K* 值的聚类算法，将聚类结果参与到 Hilbert-R 树构建当中，对矢量时空数据采用 Redis 丰富的数据结构进行分层组织存储；结合时间和聚类结果分层结构建立分级索引机制。在此基础上，进行数据密集型测试、时空范围查询及目标矢量对象查询，在真实数据集上进行，结果表明，提出的时间结合聚类分层组织存储及分级索引机制在时空查询性能方面较 CCR+树有显著提升。

1 基础建模

本节将从定义时空数据，利用 Hilbert 曲线对空间数据编码，设计空间聚类算法等方面进行阐述。

1.1 时空数据建模

时空数据是用来描述现实世界中空间实体及各实体间随时间发展变化的过程和规律。它是描述矢量数据组织和设计空间数据库的基础[11]。本文参照国际标准化组织(Open GIS Consortium，OGC)制定的简单要素规范对矢量时空数据进行组织。每个时空要素有以下重要特征：

(1) 空间特征，指实际地物的几何特征（如形状、大小和位置），以及各地物间的位置关系，通常用坐标(*x*,*y*)表示经纬度。

(2) 时间特征，通常指时空数据的空间特征和属性特征随时间而变化的，本文数据集中的时间指采集该时空要素的时刻。

(3) 属性特征，空间实体的类别和属性。

基于以上特征，对时空要素可表示为[*x*,*y*,*t*,*attrs*]，其中 *x*，*y* 分别表示要素横纵坐标，采用经纬度值代替；*t* 表示采集该要素的时间；*attrs* 表示时空要素的属性。

1.2 时空数据编码

本文对时空数据编码采用 Hilbert 曲线，其分形技术将多维空间区域等分为众多网格，并映射为一维区间，作为 Peano 族空间填充曲线家族中的分支之一，相较于 Z-Order 填充曲线，具有较优的空间聚类效果[12]。Hilbert 曲线对空间相邻近特征的区域能够做到较好地表达，即在空间联系密切且空间几何距离相近的位置上，其曲线之间的距离也会较短[13]。如图 1 显示 1、2、3 阶 Hilbert 曲线，本文选用 3 阶 Hilbert 曲线进行研究。

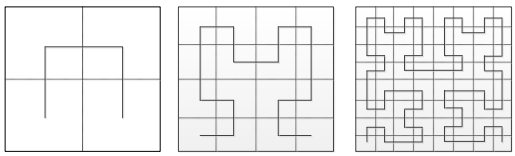

图 1 Hilbert 曲线

Fig. 1 Hilbert curves

结合矢量时空数据，对 Hilbert 曲线进行描述如下：

(1) 等分区域(cell)的大小取决于 Hilbert 曲线的阶数 λ，网格数 $m=2^{2\lambda}$；

(2) encodeNode(*x*,*y*)，利用 Hilbert 编码规则将空间 S 内的坐标(*x*,*y*)转换为 Hilbert 值。

(3) encodeSegment(x_1,y_1,x_2,y_2)，将空间区域内的坐标范围(x_1,y_1,x_2,y_2)转换为一系列 Hilbert 范围值，即一维区间。

1.3 空间聚类建模

针对时空数据空间部分采用树型空间索引易产生大量重叠和死空间，考虑将空间上相邻的数据存放于同一子树下，即与聚类技术结合的方式，在减少数据存储冗余及 I/O 寻道时间，提高检索效率等方面具有重大的研究意义。目前多数聚类算法的结果易受初始 *K* 值及离群空间数据的影响[14]，本文在对空间要素对象进行 Hilbert 曲线编码后，依据空间要素分布特点，定义距离函数和均方差准则函数。

定义 1（**距离函数**）由 $r_1,r_2,\ldots,r_n$ 组成的 *n* 个 *Rd* 空间要素集合，设 *Oj* 为第 *j* 个类的聚类中心，则 *ri* 与 *Oj* 的距离为：

$$d(r_i,o_j)=\sqrt{\sum_{l=1}^{d}(r_i^l-o_j^l)^2} \tag{1}$$

定义 2（**均方差准则函数**）为使聚类效果满足“类内紧凑， 类间分离”的原则，需要聚类测度函数进行衡量，均方差准则函数如下：

$$E=\sum_{j=1}^{k}\sum_{i\in S_j}(i-o)^2 \tag{2}$$

其中，*k* 为聚类数；*i* 为 S_j 中的数据； *o* 为 S_j 的聚类中心。算法思想：将 *k*=1 作为初始聚类，再令 *k* 值递增，每次递增后，都将通过计算得到新的聚类中心，使所有类中数据与聚类中心的距离平方误差之和最小。递增 *k* 值后，比较函数值 *E*，若该值收敛，则将 *k* 值作为聚类数；若不收敛，则继续递增 *k* 值。经分析，该算法的复杂度为 O(*nkt*)，其中 *n* 指空间要素对象的个数，*t* 指算法迭代次数。通常 *k*<<*n* 且

$t<<n$，因此该算法时间复杂度接近线性，适合处理大规模的时空数据。

2 面向 Redis 的时空索引

2.1 构建基于动态定值聚类的 Hilbert-R 树

目前，R 树及其变体的构建以空间要素逐一插入的动态过程为多，而此处的 Hilbert-R 树是静态批次构建的过程，在对空间要素进行 Hilbert 编码之后，根据定义 1 的距离函数选取两两距离最大的空间要素作为初始聚类中心，以使类间相似度较低，而类内相似度较高。递增聚类数 K 值，遵循层次聚类中的分裂思想，找到距聚类中心最远的空间对象及距该空间对象最远的另一空间对象，将二者作为新的聚类中心，开始聚类；根据定义 2 的均方差准则函数，计算并比较 K 值变化（即每次聚类）后的函数值，若该值收敛，可将当前状态视为聚类结果，否则，继续递增 K 值。对 Hilbert-R 树的构建步骤阐述如下：

(1) 以某个时刻 $T(t_0 \leqslant T \leqslant t_n)$的空间数据集为研究对象，计算该数据集下所有空间要素的最小包围矩形(Minimum Bounding Rectangle，MBR)及其中心。

(2) 调用聚类算法，产生以空间点要素及空间要素 MBR 为对象的 K 个聚类。

(3) 计算所得各聚类内包含空间要素对应的 Hilbert 值，并对该值进行升序排序。

(4) 依据节点最大容量 $M(M_{size} \leqslant 32MB)$，对排列好的空间要素构建叶节点。若聚类内空间要素数小于等于 M，则该聚类内的所有空间要素构成一个叶节点；反之，则生成多个含有 M 个空间要素的分组，并按生成顺序构建叶节点。

(5) 依据生成顺序自底向上地构建各层中间节点和根节点，进而生成基于聚类的 Hilbert-R 树。叶节点存放聚类所包含空间要素 MBR 对应的 Hilbert 值，各层中间节点则存放其子节点的 Hilbert 最大值。

基于动态定值聚类的 Hilbert-R 树构建算法如下：

```
Input: tList=(t0,tn)//按时间划分的时空数据集
Output: Hilbert-R 树
Begin
 //计算每个空间要素 MBR
 foreach element in tList
     element.MBR=cal()
 endfor
 //调用聚类算法对各空间要素聚类
 myKNN_List=myKNN(tList)//对各空间要素聚类
 //统计各聚类空间的 Hilber 值
 my_HilbertValue=statistic(myKNN_List)
 //按 Hilber 值排序
 my_SortValue=sort(my_HilbertValue)
 //按照预定的 M 值来构建树
 if value<=M then //按照排序后的 h 值构建树
  Hilbert-R=Create_tree(value,my_SortValue)
 elseif//对部分聚类构建树
//生成多个空间要素分组，并按生成顺序构建子树
  myGroup=Group(value,my_SortValue)
  Hilbert-R=Create_tree(value,myGroup)
 endif
 return Hilbert-R
end
```

2.2 建立分级索引结构

经上文 Hilbert-R 树的构建可知，Hilbert-R 树具有中间节点和叶节点两种节点结构，依据 Redis 数据库 Key-Value 方式对时空数据进行存储。

选择 Redis 作为时空数据存储数据库的原因如下：Redis 是分布式非关系型数据库之一，是一种具有高性能键值对的内存数据库。其中，键值(key)是 Redis 存储数据的重要依据，具有唯一性，同时也作为对数据检索、定位、修改和删除的根本因素[15]。Redis 中的键值对类似于关系数据库中的行和列，不同的是 value 这一列可以存储多个条目，而不是只能存储一个值[18]。Redis 支持丰富的数据结构，用于满足不同的矢量时空数据存储要求，包括：string、list、set、zset(sorted set)、hash 等多种键值数据类型。这些数据类型都支持取集合的交、并、差等多样操作，还支持不同方式的排序，并且这些操作都是原子性的。此外，Redis 为用户提供了可靠的持久化方案和主从复制功能，以确保数据的安全性。基于以上特性，Redis 可以保证对矢量时空数据的存储和管理。

本文提出一种分级索引机制，将该时空数据库结构划分为：已聚类的空间部分和时间部分。整个时空数据集被周期化的时间值（以小时为单位）所划分，即每个空间数据集对应不同的时间值，因此将时间部分的索引建立于第一层索引结构中；对某一时刻 T 的聚类结果（即 Hilbert-R 树中具有相同时间值的中间节点）进行索引信息的存储构建；而对空间要素的索引建立在 Hilbert-R 树的中间节点之上，即图 2 第三层，中间节点存放其子节点的最大 Hilbert 值。基于以上分层分级的索引结构，查询某个时空数据可以先通过时间索引，确定该时刻的空间聚类信息，再利用高效的 Hilbert-R 树定位于目标对象。具体分层分级索引结构及其在 Redis 数据库中的存储方式如图 2 所示。

第一层：时间索引信息存储，采用 Redis 数据库中的有序集 sorted set 数据结构对时空数据集元数据信息中唯一对应的时间值进行组织，可以通过 key=Index: T 的格式访问对应的数据集，将时间值[n,n-1)作为 sorted set 使用的排序码，如表 1 所示。

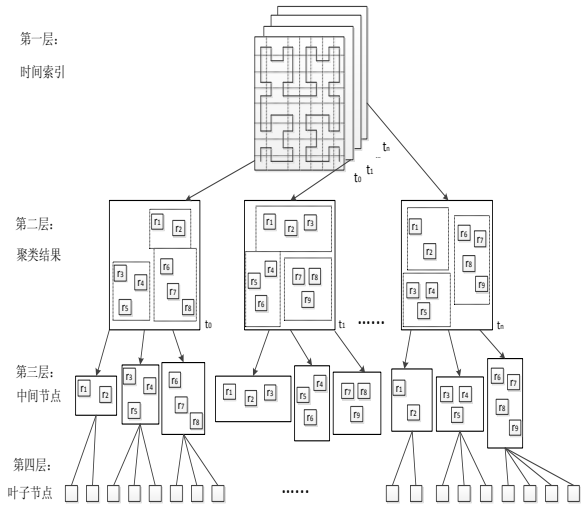

图 2 分级索引结构

Fig. 2 Hierarchical index structure

表 1 时间索引信息存储结构

Tab. 1 Time index information storage structure

Key	Value
Index：T_0	[0,1)
Index：T_1	[1,2)
…	…
Index：T_n	[n,n-1)

第二层：调用聚类函数，将某一时刻[n,n-1)下的空间要素（r_1,r_2,…,r_n）进行聚类操作，得到 K 个聚类结果。统计类中要素的 Hilbert 值，并升序排列。

第三层：中间节点索引信息存储，针对同一时间下不同聚类结果的索引采用 Redis 的 sets 数据结构进行组织，通过 key=Index：T：Index_I 可以访问具体的聚类索引项，即 Hilbert-R 树的中间节点，他们都具有相同的时间属性，具体结构如表 2 所示，I_n 数为 T_n 时刻的聚类个数。

表 2 分级索引信息存储结构

Tab. 2 Hierarchical index information storage structure

Key	Value
Index：T_0	Index：I_0
	Index：I_1
	…
	Index：I_n

表 3 中间节点信息存储结构

Tab. 3 Intermediate node information storage structure

Key	Value
Index：I_0	$Max_0(rh_1,rh_2,…,rh_n)$
Index：I_1	$Max_1(rh_1,rh_2,…,rh_n)$
…	…
Index：I_n	$Max_n(rh_1,rh_2,…,rh_n)$

第四层：Hilbert-R 树的叶子节点，即各聚类内所包含的空间要素（r_1,r_2,…,r_n）。

最后将 Hilbert-R 树的中间节点元数据信息利用 Redis 数据库中的 sets 数据结构进行组织，用于存放叶子节点的最大 Hilbert 值。如表 3 所示，其中（rh_1,rh_2,…,rh_n）表示各叶子节点所对应的 Hilbert 值。

2.3 基于 Hilbert-R 树的时空查询流程

时空范围查询的基本思想：根据查询空间范围(x,y)或(x_1, y_1,x_2,y_2)调用函数 encodeNode(x,y)或 encodeSegment (x_1,y_1,x_2, y_2)并转换为对应 Hilbert 值或 Hilbert 范围值；同时结合所查询时间值调用第一级时间索引，确定查询目标所对应的 Hilbert-R 树中，递归调用 Hilbert-R 树索引结构，对树中空间要素与查询对象进行精准相交检测，直到输出查询结果。图 3 是时空查询算法流程图。

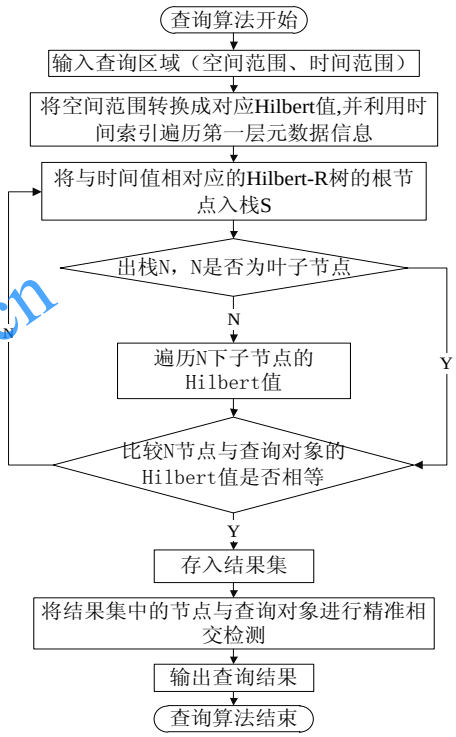

图 3 时空查询算法流程图

Fig. 3 Flow chart of spatial-temporal query

3 实验分析

实验采用 Redis 2.6 标准版，操作系统为 CentOS release 6.5 64 bit，硬件环境配置为：Intel(R) Core(TM) i7-4790 CPU @ 3.60GHz(3601 MHz)，8G 内存，500G 硬盘。实验数据集采用 ShapeFile 格式的乌鲁木齐市道路交通数据，选取了 2011 年-2014 年的数据。考虑时空范围和空间分布两个因素，分别进行了数据密集型测试及目标矢量对象测试，最后进行时空范围查询综合测试，实验对比文献[9]中的 CCR+树，并对本文方案技术特点进行验证如下。

3.1 数据密集型测试

本实验选取相同时空范围内不同空间分布状况的数据集对 Hilbert-R 树及文献[9]CCR+树对空间索引性能的测试。

在定义 1 的基础上，**定义 P 为平均分布状态**[6]

$$P = \frac{1}{n}\sum_{i=1}^{n} 1_i \qquad (3)$$

用 $l_i(i=1,2,\ldots,n)$表示空间数据的邻近对象个数，P 值的大小反映了空间分布的密集状态。如表 4 所示数据集 T1，T2 的 P 值较大，则说明空间分布状态较密集，而数据集 T3，T4 的 P 值接近 1，表示二者空间分布较均匀。

表 4 空间数据分布状态

Tab. 4 Distribution of spatial data

数据集	P 值
T1	3.44
T2	3.10
T3	1.07
T4	1.05

不同空间分布下，Hilbert-R 树与 CCR+树的查询耗时结果如图 4 所示。在不同密集程度下，经聚类后的 Hilbert-R 树的查询耗时约为 CCR+树的一半，说明前者查询耗时较短，进一步验证了数据是否平均分布对本文算法的查询效率影响不大。这是因为本文算法充分考虑了空间要素的分布特点，动态确定 K 值的聚类算法尽可能地将空间分布相近的要素聚为一类，经有效聚类后构建 Hilbert-R 树，提出分层分级的索引结构，将聚类结果在同一子空间的要素组织在同一子树下，较大程度上降低中间节点间的重叠，对分布不均匀的空间要素得以正确处理，从而提高索引操作的性能。

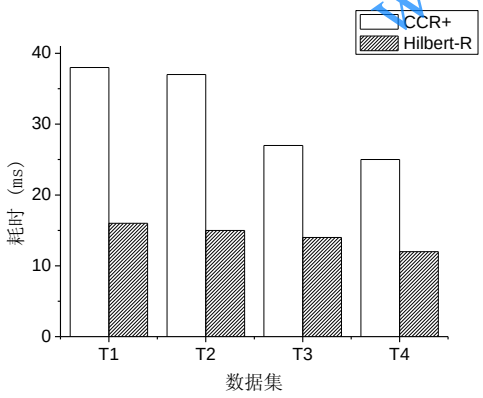

图 4 不同分布状态下 CCR+和 Hilbert-R 的查询对比

Fig. 4 Contrast of CCR+ and Hilbert-R in different distribution states

3.2 目标矢量对象查询

为进一步分析空间分布相同（分别选取分布密集和均匀的情况下）不同时空范围下针对目标矢量对象的查询，本实验数据采用时空数据生成器（Generator of Spatio Temporal Data，GSTD）生成不同的时空范围模拟数据集，其中 D1-D5 的 P 值都等于 3.44，D6-D10 的 P 值都等于 1.05，分别选取不同时空范围的区域进行研究，对应矢量对象数目也不尽相同（如表 5 所示）。GSTD 是一个被广泛认同的目标对象数据集生成器。利用所生成的诸模拟数据集分别查询目标矢量对象，计算各自访问节点数目的平均值，以此反映 Hilbert-R 树与 CCR+树的查询效率。

表 5 样本数据集描述

Tab. 5 Sample data set description

时间范围(h)	空间范围(km)	矢量对象数目	
		密集	均匀
1	1	D1：6906	D6：4850
2	2	D2：10267	D7：8059
3	3	D3：15370	D8：13168
10	10	D4：43579	D9：39080
20	20	D5：83650	D10：75106

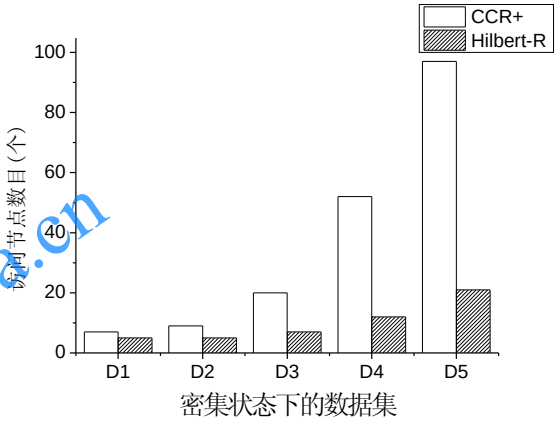

图 5 密集分布下的访问节点数对比

Fig. 5 Contrast of the number of access nodes under dense distribution

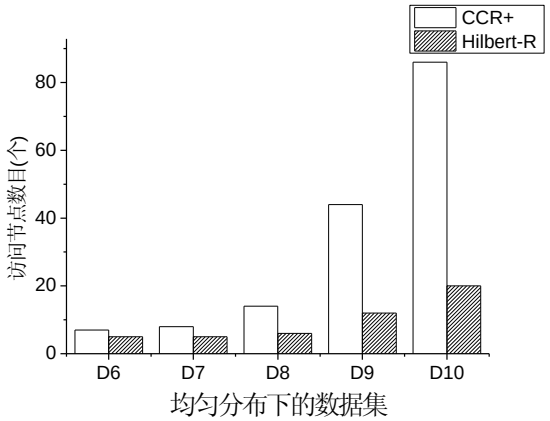

图 6 均匀分布下的访问节点数对比

Fig. 6 Contrast of the number of access nodes under uniform distribution

图 5 和图 6 分别展示了密集分布和均匀分布下，Hilbert-R 树与 CCR+树对目标矢量对象查询时，二者平均访问节点数目的情况：

1）CCR+树访问节点数与数据集的数量呈正比例增长的趋势；而 Hilbert-R 树无论在分布密集还是均匀的情况下，对节点访问数的增幅都不明显；

2）随数据集的增大，Hilbert-R 树对节点访问的变化率较 CCR+树稳定，从而验证了本算法查询效率对数据量的变化不敏感，受数据分布状态的影响不大。

这是因为 CCR+树进行查询时从索引树的根节点开始，依次遍历存储于其中的空间要素，节点间被动搜索满足条件的叶节点，有可能会遍历整棵树。

本文在构建 Hilbert-R 树的过程中，充分考虑实际时空数据的分布特点，利用聚类方法使得 Hilbert-R 树结构更加紧凑。算法动态寻找最优 K 值将数据对象划分为多个类，自底向上逐层构建树结构，确保数据精准划分，迭代重定位使划分逐步改进，同时分层分级索引结构的提出对时间维的查询更加便捷，避免访问多余节点，有利于需要同时考虑空间和时间的目标矢量对象的查询。

3.3 时空范围查询

本实验选取了 2013 年 2 月 1 日-2 日的车辆轨迹数据（其中共包含 30177 条记录数）作为实验数据对聚类的 Hilbert-R 树及 CCR+树进行时空范围查询的综合测试，选中从 5084 条记录数开始，以平均间隔约 5000 条递增的数据量（包括：5084、10850、15065、20037、30177 条记录数）分别进行 10 次测试后，取查询耗时的平均值。

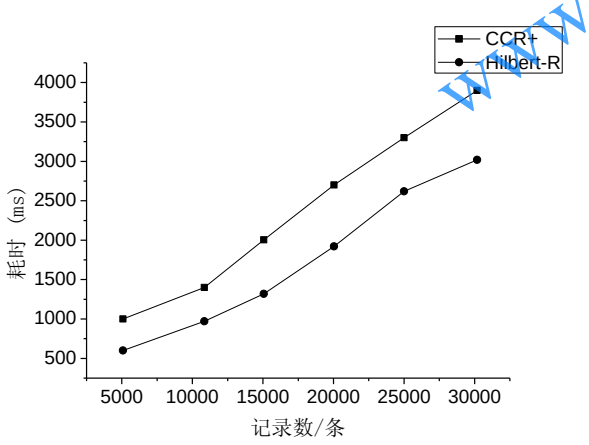

图 7 时空范围查询结果

Fig. 7 Results of spatio-temporal range queries

图 7 对比了聚类的 Hilbert-R 树与 CCR+树在不同数据量下的时空范围查询效率，可分析得到：

1）Hilbert-R 树的查询耗时与数据记录数量的关系呈线性正相关；

2）Hilbert-R 树相比 CCR+树的查询耗时平均缩短了约 25%；

3）随着数据量的增大，Hilbert-R 树的查询响应时间优势较 CCR+树更趋明显。

经分析发现，在对时间范围方面的查询时，由于 CCR+树并未考虑时空数据的时间属性，将其归入空间属性的另一维，致使整个多维空间数据不能协调分布，节点间频繁访问，徒增查询开销。

然而 Hilbert 曲线本身具有良好的数据聚集特性。聚类后的 Hilbert-R 树的构建对不同分布的空间要素进行处理，使相邻空间要素聚集于相邻中间节点之上。降低了 R 树节点间的重叠，使得查询所需访问的节点数更少，提高了查询效率。同时本文提出的分层分级索引结构，充分考虑时间属性，时空数据依时间-空间合理组织，对于查询性能的提升也起到了重要作用。

3.4 小结

为了探索在不同空间分布及时空范围下，Hilbert-R 树及 CCR+树对矢量时空数据的查询情况，本节采用 ShapeFile 格式的乌鲁木齐市道路交通数据展开实验，分别对 Hilbert-R 树及 CCR+树的数据密集型、时空查询效率和目标矢量对象查询进行了测试。分析实验结果可得出结论如下：

1）聚类后的 Hilbert-R 树分层分级索引结构受数据分布状况的影响不大，因此适用于对空间分布不均且局部密度较大的时空数据查询；

2）随数据量的增加，Hilbert-R 树相较于 CCR+树的时空查询效率更高，其访问节点数更少，因此适用于时空范围较大的矢量数据。

综上，采用聚类的 Hilbert-R 树分层分级索引结构可开展时空大数据存储及查询算法的应用，在内存数据库支持下进行多尺度，多类型的时空数据查询，为时空 GIS 的发展奠定理论基础及方法借鉴。

4 结语

鉴于时空数据的矢量查询应用中缺乏对时间维和空间要素合理组织的考虑，本文从时空数据的空间分布出发，提出基于Hilbert-R 树分级索引的时空查询算法，该算法利用 Redis 丰富的数据结构对周期化的数据集建立时间索引，经 Hilbert 曲线划分并编码空间部分，在此基础上进行聚类处理。将空间聚类结果与时空数据的时间属性结合，建立 Hilbert-R 树的分层分级索引结构。经实验验证，所提算法有效降低了时空查询的时间开销，且对树节点的访问量不大，查询耗时受不同数据分布密度的影响较小，更适合 Redis 对海量时空数据高效查询处理的需求。

下一步研究工作将着眼于：(1)优化 Hilbert-R 树索引构建；(2)引入机器学习算法对时空轨迹模式挖掘进行研究，使时空查询效更优。

参考文献

[1] He Z, Kraak M J, Huisman O, et al. Parallel indexing technique for spatio-temporal data[J].Isprs Journal of Photo grammetry & Remote Sensing,2013,78(4):116-128.

[2] Huang J, Wei P, Yu H, et al. A Graph Based Bi-level Index for Spatio-temporal Data Analysis with Map Reduce[C]//Sixth International Symposium on Computational Intelligence and Design.IEEE,2014:339-342.

[3] Du N, Zhan J, Zhao M, et al.Spatio-Temporal Data Index Model of Moving Objects on Fixed Networks Using HBase[C]//IEEE International Conference on Computational Intelligence & Communication Technology.IEEE,2015:247-251.

[4] Zheng L, Zhou L, Zhao X, et al. The Spatio-Temporal Data Modeling and Application Based on Graph Database[C]// International Conference on Information Science and Control Engineering. IEEE Computer Society, 2017:741-746.

[5] Wang J.Optimization algorithm for R-tree combining with spatial-clusting[J].Computer Engineering & Applications,2014,50(5):112-115.

[6] 胡昱璞,牛保宁,HUYupu,等.动态确定K值聚类算法的R-树空间索引构建[J].计算机科学与探索,2016,10(2):173-181.(HU Y P,NIU B N.R-tree Spatial Index Construction Based on Dynamical K-value Clustering Algorithm [J].Journal of Frontiers of Computer Science and Technology,2016,10(2):173-181.)

[7] 李松,崔环宇,张丽平,等.基于 CURE 聚类算法的静态 R 树构建方法[J].计算机科学,2015,42(10):193-197.(LI S,CUI H Y,ZHANG L P, et al. Static R-tree Building Method Based on Cure Clustering Algorithm[J].Computer Science,2015,42(10):193-197.)

[8] 汪璟玢.一种结合空间聚类算法的 R 树优化算法[J]. 计算机工程与应用,2014,50(5):112-115.(WANG J. Optimization algorithm for R-tree combining with spatial-clusting[J].Computer Engineering and Applications,2014,50(5):112-115.)

[9] 戚将辉,张丰,杜震洪,等.基于内存数据库的矢量数据存储与空间索引研究[J].浙江大学学报(理学版),2015,42(3):365-370.(QI J H,ZHNAG F,DU Z H, et al. Research of the land use vector data storage and spatial index based on the main memory database. Journal of Zhejiang University(Science Edition),2015,42(3):365-370)

[10] 罗敬宁,刘立葳.遥感大数据分布式技术研究与实现[J].应用气象学报,2017,28(5):621-631.(LUO J N,LIU L W. Research and Implementation of Remote Sensing Big Data Distributed Technology[J].Journal of Applied Meteorological Science,2017,28(5):621-631.)

[11] 张翀, 陈晓莹, 史宗麟,等. HBase 时空查询算法研究[J].小型微型计算机系统,2016,37(11):2409-2415.(ZHANG C，CHEN X Y，Shi Z L,et al. Algorithms for Spatio-temporal Queries in HBase[J].Journal of Chinese Computer Systems,2016,37(11):2409-2415.)

[12] Kim H I, Hong S, Chang J W. Hilbert curve-based cryptographic transformation scheme for spatial query processing on outsourced private data[M].Elsevier Science Publishers B.V.2016,104:32-44.

[13] 田丰,桂小林,张学军,等.基于兴趣点分布的外包空间数据隐私保护方法[J].计算机学报,2014,37(1):123-138.(TIAN F,GUI X L,ZHANG X J, et al. Privacy -Preserving Approach for Out sourced Spatial Data Based on POI Distribution[J].Chinese Journal of Computers.2014,37(1):123-138.)

[14] Kumar K M, Reddy A R M.An Efficient k-Means Clustering Filtering Algorithm Using Density Based Initial Cluster Centers[J].Information Sciences,2017,418-419(2017):286-301.

[15] 丁建立, 郑峰弓, 李永华,等. 基于 NoSQL 的海量航空物流小文件分布式多级存储方法[J].计算机应用研究,2017,34(5):1433-1436.(DING J L,ZHENG F G,LI Y H,et al.Method of distributed multi-level storage of massive small files of air logistics based on NoSQL[J].Application Research of Computers,2017,34(5):1433-1436.)

[16] 焦健,李岩.基于Redis的SVG空间信息可视化数据库[J].小型微型计算机系统,2015,36(6):1193-1198.(JIAO J，LI Y.SVG Spatial Visualization Database Based on Redis[J].Journal of Chinese Computer Systems,2015,36(06):1193-1198.)

[17] Kabakus A T,Kara R.A performance evaluation of in-memory databases[J].Journal of King Saud University-Computer and Information Sciences,2016,29(4):520-525.

[18] 朱进,胡斌,邵华,等.基于内存数据库Redis的轻量级矢量地理数据组织[J].地球信息科学学报,2014,16(2):165-172.(ZHU J, Hu B, SHAO H,et al. Research of Lightweight Vector Geographic Data Management Based on Main Memory Database Redis[J].Journal of Geo-information Science,2014,16(02):165-172.)

作者简介：

基金项目：国家自然科学基金资助项目(61562086，61462079)；新疆维吾尔自治区教育厅项目(XJEDU2016S035)；新疆大学博士科研启动基金项目(BS150257)；新疆维吾尔自治区教育厅创新团队(XJEDU2017T002)

侯海耀(1990-)，男，山西省汾阳人，硕士研究生，主要研究方向：时空数据库索引和遥感图像处理；通讯作者：钱育蓉(1980-)，女，山东省武城人，博士，教授，CCF 高级会员(23806S)，主要研究方向：网络计算和遥感图像处理；英昌甜(1989-)，女，新疆乌鲁木齐人，博士，主要研究方向：图像处理、 内存计算；张晗(1987-)，男，辽宁省本溪人，硕士研究生，研究方向：图像处理与模式识别；卢学远(1992-)，

男，浙江温州人，硕士研究生，CCF 会员，主要研究方向：图像处理、机器学习；赵燚(1993-)，女，新疆克拉玛依人，硕士研究生，主要研究方向：时空数据库索引、遥感图像处理。

This work is partially supported by the National Natural Science Foundation of China (61562086，61462079),the Funds for Education Department Project of Xinjiang Uygur Autonomous Region(XJEDU2016S035), Doctoral Research Startup Fund Project of Xinjiang University(BS150257) , the Funds for Creative Research Groups of Higher Education of Xinjiang Uygur Autonomous Region(XJEDU2017T002)

HOU Haiyao, born in 1990, M. S. candidate . His research interests include spatiotemporal database index, remote sensing image processing.

QIAN Yurong ,born in 1980, Ph.D., Professor, Senior Member of CCF (23806S). Her research interests include network computing and remote sensing image processing.

YING Changtian,born in 1989,Ph.D.Her research interests include image processing ,in-memory computing.

ZHANG Han, born in 1987, M. S. candidate . His research interests include image processing and pattern recognition

LU Xueyuan, born in 1992, M. S. candidate . His research interests include image processing , machine learning.

ZHAO Yi, born in 1993, M. S. candidate . Her research interests include spatiotemporal database index, remote sensing image processing.