



# 地理格网模型支持下的轨迹数据管理与分析框架:方法与应用

李 军<sup>1</sup> 刘举庆<sup>1</sup> 赵学胜<sup>1</sup> 黄 骞<sup>2</sup> 孙文彬<sup>1</sup> 许志华<sup>1</sup> 王 昊<sup>2</sup>

1 中国矿业大学(北京)地球科学与测绘工程学院,北京,100083

2 华为技术有限公司服务实验室,北京,100095

**摘要:**近年来,各类位置感知设备产生的轨迹数据被广泛应用于城市规划、智能交通、公共卫生、行为分析等各个领域,但是常规矢量表达方式及建立在其基础之上的分析算法的计算复杂度高,无法满足大规模轨迹数据的高时效性应用需求。针对上述问题,提出了基于地理格网模型的轨迹数据管理与分析框架,为轨迹数据挖掘的“表达-管理-分析-应用”全链条研究提供新的技术框架,主要包含地理格网模型、轨迹多尺度表达与组织、轨迹计算与分析、高性能计算技术、轨迹挖掘应用5部分。介绍了各部分的实现思路和方法,并阐述了格网模型与轨迹数据结合的优势,包括存储管理高效灵活、适合高性能计算技术和契合自动控制与智能计算需求等。以城市交通流多层次实时可视化和基于地理格网编码的相似性分析两个应用实例验证了该技术框架理论与技术方法的可靠性和有效性。

**关键词:**轨迹管理;轨迹分析;地理格网模型;移动对象;数据挖掘

**中图分类号:**P208 **文献标志码:**A

自卫星定位、传感网络、无线通信等技术出现与普及以来,各个领域产生了大量轨迹数据,并正在持续累积<sup>[1]</sup>。轨迹数据是指一个或多个移动对象在运动过程中由定位设备在不同时刻采集的地理位置、运动状态及相关属性的数据,用数学公式通常可表示为按时间顺序排列的一串轨迹点集合 $T = \{p_1, p_2, p_3 \cdots p_n\}$ ,其中 $p_i = \{x, y, a, t\}$ , $(x, y)$ 是轨迹点坐标, $a$ 是轨迹点的属性值, $t$ 是信息采集时间。轨迹数据来源多种多样,可以通过全球卫星导航定位、移动通信基站定位、射频识别技术、摄像头视觉定位、卫星遥感影像识别和手机端应用等不同方式获取。

按照移动对象的类型,轨迹数据主要包括人类活动轨迹、交通车辆轨迹、动物活动轨迹和自然现象轨迹4种类型<sup>[2]</sup>。轨迹数据能够反映移动对象个体或群体在精细尺度上的位置及状态的全过程变化,具有丰富的时间、空间、语义及交互信息特征,因此具有很大的研究价值<sup>[3]</sup>。轨迹数据挖掘已广泛应用于智能交通<sup>[4-5]</sup>、城市规划<sup>[6-8]</sup>、公共安全<sup>[9-10]</sup>、环境保护<sup>[11]</sup>、社会行为<sup>[12-13]</sup>、文化旅

游<sup>[14-15]</sup>等领域。

为了尽可能地挖掘轨迹数据蕴含的应用价值,学者们在轨迹数据管理与分析方法方面开展了大量研究工作<sup>[16-17]</sup>,其中轨迹数据管理作为数据挖掘与应用的基础,一直以来是国内外研究的热点,主要包括轨迹数据表达模型、轨迹数据索引与检索、轨迹数据库、轨迹数据预处理(数据清洗、数据压缩、轨迹分段、路网匹配)等几个方面<sup>[1]</sup>。另外,在轨迹数据分析方面也产生了大量的研究成果<sup>[2]</sup>,如为研究移动对象的代表性路径或公共倾向行为的轨迹聚类<sup>[18]</sup>、用来识别轨迹的不同运动状态以区分出行方式和活动类型的轨迹分类<sup>[19]</sup>、研究移动对象群体行为特征和规律的模式挖掘,包括伴随模式挖掘<sup>[20]</sup>、频繁模式挖掘<sup>[21]</sup>、周期模式挖掘<sup>[22-23]</sup>、异常行为探测<sup>[24]</sup>等。

但进入大数据时代以来,轨迹数据呈现井喷式增长,轨迹大数据的“5V”特性为其挖掘分析带来了巨大困难<sup>[25]</sup>,学者们尝试将大数据处理技术应用于轨迹研究,即利用云计算、分布式计算、并行计算等高性能计算模式加速大规模轨迹数据

收稿日期:2020-08-31

项目资助:国家自然科学基金(41971355);中国矿业大学(北京)越崎青年学者资助计划。

第一作者:李军,博士,副教授,主要研究方向为地理数据分析与挖掘。junli@cumtb.edu.cn

通讯作者:刘举庆,博士生。juq.liu@ student.cumtb.edu.cn

分析<sup>[26-29]</sup>,但此类方法主要依赖计算资源,不是从算法本身解决轨迹数据计算与分析复杂的问题。也有部分学者探索优化索引结构或提高算法性能,如文献[17]和文献[30]分别提出了Geohash-Tree和TripCube轨迹数据索引结构实现对轨迹数据的高效管理,文献[31]和文献[32]分别对相似性分析与地图匹配算法进行优化,降低算法复杂度,但算法效率未得到数量级提高,应对大规模轨迹数据仍然存在困难。

综上所述,虽然现有研究工作取得了一些积极的成果,但是现有轨迹管理与分析方法在同时面对大规模、复杂异构、多维动态的轨迹数据特点与高时效性应用需求时,仍然面临以下挑战:

1) 轨迹数据来源不一导致位置表达方式、空间参考、定位精度等高度不一致,具有多源异构特性,使得多源异构轨迹数据的统一管理与融合分析困难;

2) 缺乏适合于大规模轨迹数据的高效索引与快速检索方法;

3) 已有的轨迹数据分析算法计算复杂度高,难以满足即时分析的需求;

4) 矢量表达轨迹数据精度高,但跨尺度分析相对独立,无法根据不同应用场景自适应、灵活地调整轨迹数据表达尺度,以降低上层分析复杂度。

总之,现有轨迹数据管理与分析理论难以满足日益迫切的时效性高、灵活性强的轨迹数据应用需求。

近些年来,随着计算机技术、地理信息技术的快速发展,以及地理空间信息的表达、更新与分析日益复杂,地理格网空间数据模型得到重新认识和重视,并成为重要的前沿方向<sup>[33-34]</sup>。地理格网模型是由一系列离散而规则的单元,按照一定的规则组合而形成的对地理实体的表达体系<sup>[35]</sup>。它具有统一性、离散化、多尺度、多分辨率以及编码计算等特性,在多源异构数据融合、大规模空间数据高性能运算等方面具有得天独厚的天然优势<sup>[36-38]</sup>,为大规模轨迹数据管理与分析打开了新的视角。

在这样的背景下,本文提出了基于地理格网模型的轨迹数据管理与分析框架,该框架以地理格网模型为基础,实现对轨迹数据的多尺度编码表达与组织,通过表达方法、算法实现、方法策略等多方面改造并构建基于格网编码的轨迹计算与分析算法集,进一步结合高性能计算技术,为

海量轨迹数据挖掘应用提供支撑。

## 1 基于地理格网模型的轨迹数据管理与分析

### 1.1 总体框架

基于地理格网模型的轨迹数据管理与分析框架主要包含地理格网模型、轨迹多尺度表达与组织、轨迹计算与分析、高性能计算技术、轨迹挖掘应用5部分内容,涵盖了轨迹数据挖掘研究“表达-管理-分析-应用”的全链条。基本内容如图1所示,首先地理格网模型主要提供格网剖分与编码理论,以及编码转换、空间操作等基础算子,是改造轨迹数据表达、管理、分析及应用的基础理论;其次,轨迹多尺度表达与组织是在地理格网模型基础上将出租车轨迹、公交集成电路卡(integrated circuit card, IC)消费数据以及飞机自动广播相关监视数据(automatic dependent surveillance-broadcast, ADS-B)等轨迹数据映射为一维格网地址编码,并构建高效索引结构,是轨迹计算与分析的支撑技术;然后,轨迹计算与分析部分包含一系列基于格网编码空间的低复杂度、高性能计算与分析算法;而轨迹挖掘应用则针对各个领域的具体需求,提供高时效、智能的解决方案,以更好地满足实际所需,发挥轨迹数据的潜在价值;最后,高性能计算部分则借助大数据处理技术为整个框架提供高效的存储和计算资源,保证框架的高速平稳运作。

### 1.2 关键理论与方法

#### 1.2.1 地理格网模型

地理格网模型的核心内容是格网剖分与编码,是将地球表面空间按照一定的规则逐级划分,并通过地址编码的方式表达任意空间位置。地理格网模型有很多种,以二叉树格网为例,它是最流行的格网模型之一,具有矢量坐标兼容性好、索引成熟、空间分析简单等优势<sup>[39-41]</sup>,其基本原理如图2(a)所示。

首先,将地球表面空间范围逐级等分为4个相同大小的空间子区,格网单元总数为 $2^N \times 2^N$ ,其中 $N$ 为剖分层级,层级越大,空间分辨率越高,表示的空间位置越精确。然后,基于 $Z$ 阶空间填充曲线对格网单元位置进行编码,基本过程是根据剖分层级 $N$ 及其对应的格网空间分辨率计算格网单元(简称“格元”)所在的行列号 $I$ 和 $J$ ,并通过莫顿交叉编码获得格元编码。在需要考虑时间维度信息时,同样可以构造时空格网模型,即利用八

叉树和三维Z阶曲线进行剖分与时空编码(图2(b)),实现三维时空的降维表达;在需要时空分离

的应用场景中,可以解码为空间码和时间码,以满足定制化的应用需求。

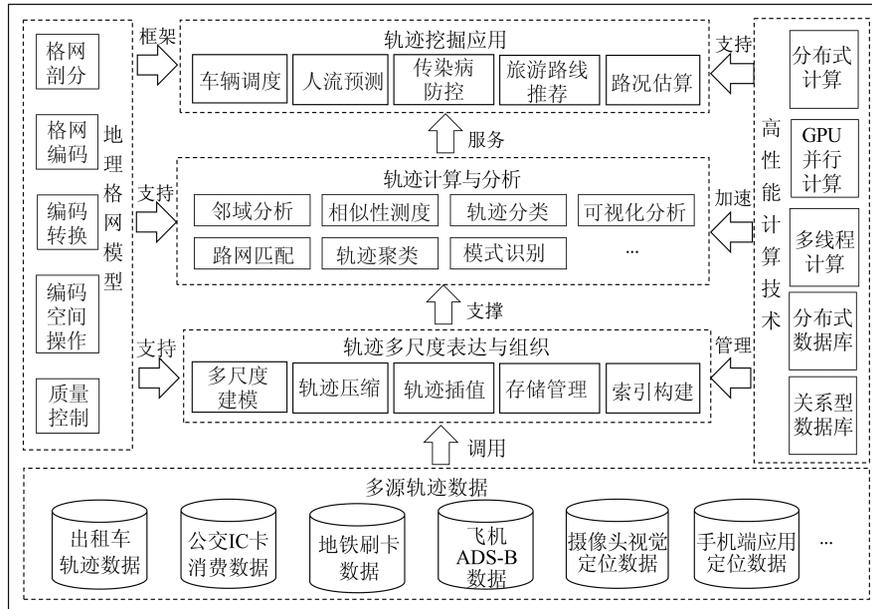


图1 基于地理格网模型的轨迹数据管理与分析框架

Fig.1 Technology Framework of Trajectory Data Management and Analysis Based on Geographical Grid Model

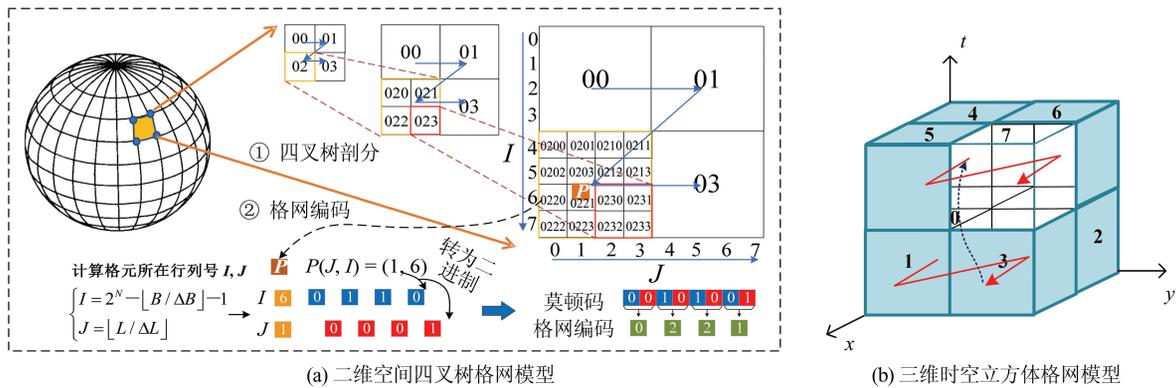


图2 地理格网剖分与编码

Fig.2 Geographical Grid Division and Coding

地理格网模型还包括编码转换、空间操作和质量控制等内容,其中编码转换是实现各类坐标参考系与格网编码间的相互转换,并可以进一步考虑引入到定位导航服务中,实现数据采集端的编码转换与输出;空间操作包含格网空间关系计算(邻近编码查询、包含关系判断等)与格网度量计算(距离、面积、方向等)两类基础算子;而质量控制是面向不同空间尺度的格网自适应剖分与编码,以实现不同类型、不同空间尺度下矢量要素格网编码表达精度与数量的平衡,并根据数据精度控制矢栅转换的误差,以满足具体应用场景下的精度要求。

### 1.2.2 轨迹多尺度编码表达与组织

经典矢量坐标只能在单一尺度无差别地表

达“精确”位置信息,往往会忽略多样化的位置感知设备带来的表达精度及采样率等方面的差异,这就使不同空间尺度场景下轨迹数据的分析与应用变得复杂且低效,例如全球尺度下高频采样的候鸟迁徙轨迹数据,在矢量坐标表达下进行相似性分析或聚类分析会增加大量计算频次。而轨迹多尺度编码表达中的多尺度建模则是在四叉树格网模型基础上,根据空间位置精度需求将轨迹数据映射到不同空间分辨率的格网层级中(见图3),支持按需调用和灵活切换,以达到表达精度和数据量的平衡。由图3可以看出,轨迹要素映射的格网层级越高,轨迹形态刻画越精细,轨迹编码数据量也会越大。另外,当采样间隔较大的轨迹要素映射到精细格网层级时,会导致轨

迹格元间不连续,则利用式(1)进行轨迹点插值(如图3中的 $P_5$ 、 $P_6$ 点),并获取对应格网编码;而轨迹采样密集时,合适的格网表达层级则会压缩轨迹数据,即同一个格网内的轨迹点合并,在满足精度的前提下减小数据冗余。

$$\begin{cases} X = X_{\text{Cell centerline}} \\ Y = \frac{Y_2 - Y_1}{X_2 - X_1} (X - X_1) + Y_1 \end{cases} \quad (1)$$

式中, $X_{\text{Cell centerline}}$ 为格元中心线的经度; $(X_1, Y_1)$ 、 $(X_2, Y_2)$ 为该轨迹上相邻两点的坐标。

轨迹的多尺度编码表达为大规模数据的存储管理和索引构建提供了极大的便利。一方面,简单的一维轨迹格网编码支持当下各种成熟的B-Tree、Hash等索引结构,并且同一个父级格网的轨迹数据拥有共同的编码前缀且表达上连续,便于使用现有数据库技术存储轨迹编码数据;另一方面,能够轻松建立轨迹要素与位置格网单元之间的对应关系,以实现二者之间的快速查询分析,尤其是快速获取经过某个格网单元的轨迹要素;此外,轨迹数据的属性信息以字段形式存储表达,便于通过轨迹编码联动分析。

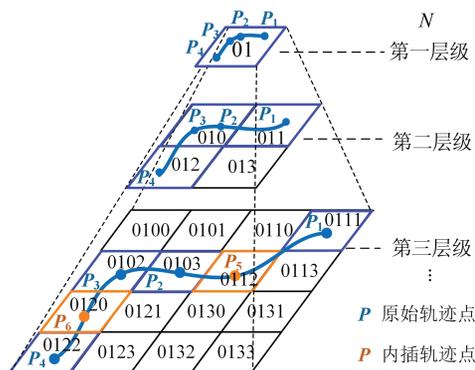


图3 轨迹多尺度编码表达

Fig.3 Multi-scale Coding of Trajectories

### 1.2.3 基于格网编码的轨迹计算与分析

轨迹数据表达模型的转变必然引发上层轨迹分析算法的革新,需要充分结合轨迹编码表达特点来探索一套面向地理格网模型的计算与分析方法,主要从以下几方面进行改进:(1)利用轨迹编码蕴含的索引信息实现轨迹的快速查询,如邻域查询、路网匹配等,基本思想是将轨迹编码作为数据索引,使得轨迹查询操作仅在索引层执行,并结合同一格网单元内轨迹编码前缀相同且时空上邻近这一原理,实现海量轨迹数据的快速时空查询。(2)利用编码运算复杂度低的优势改造算法,例如轨迹聚类、轨迹分类或相似性分析

等算法的一个重要环节是轨迹距离的度量,而经典的轨迹距离度量算法(如动态时间扭曲法(dynamic time warping, DTW)、最长公共子序列(longest common subsequence, LCSS)、实序列编辑距离(edit distance on real sequence, EDR)、豪斯多夫距离(Hausdorff distance, HD)等)的复杂度为 $O(N \times M)$ ,计算较为耗时,而基于格网编码的轨迹距离可通过辛普森距离、杰卡德距离等集合距离进行度量,使算法复杂度降为 $O(N)$ ,进而加速聚类分析等高级算法。(3)轨迹多尺度编码表达特性可以服务于可视化分析,可依据其多尺度特性在有限的表达数据能力基础上尽可能多地展现轨迹的关键分布信息,通过建立类似于栅格金字塔的结构,根据终端视域大小筛选相应尺度下的轨迹数据进行可视化,实现大规模轨迹数据自适应筛选、流畅加载与渲染。

### 1.2.4 高性能计算技术

大数据时代下,海量轨迹数据的挖掘往往需要并行式、分布式等高性能计算框架的支撑,地理格网模型自身离散性的时空表达特点可以很好地与之融合,实现轨迹大数据的分布式存储与并发式计算。首先,基于轨迹编码数据蕴含的尺度信息构建“粗-细”粒度协同的多级时空索引机制(见图4),减少数据读写的总量,以缩减响应时间,提高大规模轨迹数据的组织管理和查询效率。即根据轨迹数据来源分为多个数据集,然后利用大尺度地理格网将数据分块,以粗粒度格网编码作为数据块的索引信息;进而利用细粒度格网编码构建时空索引,由于轨迹编码的一维特性,时空索引可以采用成熟且通用的B-Tree、Hash等索引结构,不仅可以在关系型数据库中应用,同样适用于分布式数据库。在此基础上,格网划分下的块状轨迹数据可以自动分配到Hadoop、Spark等分布式或图形处理器(graphics processing unit, GPU)并行计算框架下的各个节点,并且分布式计算的架构有助于多级索引机制的实现,为海量数据的并发式查询和挖掘提供了双向动力。

### 1.2.5 轨迹挖掘应用

轨迹挖掘相关应用众多,但往往都需要周期性或实时性的服务,对时间效率要求高,而基于地理格网模型的轨迹数据管理与分析框架为满足这些实时或准实时的应用需求提供了可能。例如在智能决策领域,面对大规模出租车实时调度问题,以格网快速汇聚行人打车需求,并进行

邻近车辆即时搜索,为出租车智能调度及峰时计价提供解决方案。智能推荐系统往往较为复杂,需要融合多维信息来综合分析判断,如旅游路线的推荐,但轨迹格网编码可以将多源信息聚合,建立精细而丰富的轨迹特征库,以便轨迹精确分类并进行同类旅游路线推荐。针对易传播、难阻断、致死率高的流行病肆虐问题,以个体出行轨迹为数据基础,通过基于格网编码的时空邻域查询、接触分析算法,在海量人群中对流行病接触者进行快速识别与追踪,阻断病毒传播链,为城市传染病防控提供技术支持。以格网为感知单元可以实现城市的动态监测与管理,如实时人流的感知与预测,其核心在于人流的高效查询与运动规律的挖掘,并可通过格网实现轨迹多尺度可视化。轨迹挖掘中的其他应用如路况估算、道路图更新、航空空域资源精细管理等也可以在此框架下实施,以更好地服务于城市管理、智能交通、公共安全等领域。

算复杂度,尤其是对于存在复杂几何运算和大量浮点型计算的轨迹分析算法进行改造,更能体现地理格网与轨迹数据分析结合的巨大潜力,如轨迹相似性计算、轨迹聚类、轨迹分类以及异常或伴随轨迹识别等。由以上分析可以得知,基于地理格网模型的轨迹数据管理与分析框架涵盖了绝大部分的轨迹计算与分析方法,均可以通过编码计算改造提高计算效率,但对于拓扑关系及网络分析的改造较为困难。

2) 适合于高性能计算技术。高性能计算是未来大规模轨迹数据挖掘的一个重要方向,而格网编码表达下的轨迹数据天然以空间格网单元分割,具有块状离散特性,有利于轨迹大数据的分布式存储与管理,适合分布式存储、云计算、GPU并行计算等高性能处理技术。而且格网编码可以直接采用二进制存储与计算,与计算机集成电路的存储机制相匹配,能够将算法集成到芯片中,实现了从计算模式和硬件层面上的双提速,为高时效的轨迹挖掘应用提供了技术支持。

3) 契合自动控制与智能计算需求。精细的格网空间能为移动对象的自动控制提供支撑,为其全过程提供统一的定位、表达、感知和操控框架,如无人驾驶车辆、机器人等。首先,精细格网提供高精度的位置信息;然后,以统一的格网编码为载体进行时空信息表达,并基于分析算法实现周边环境动态感知和运动特征建模,进而实现全方位的精细化控制。

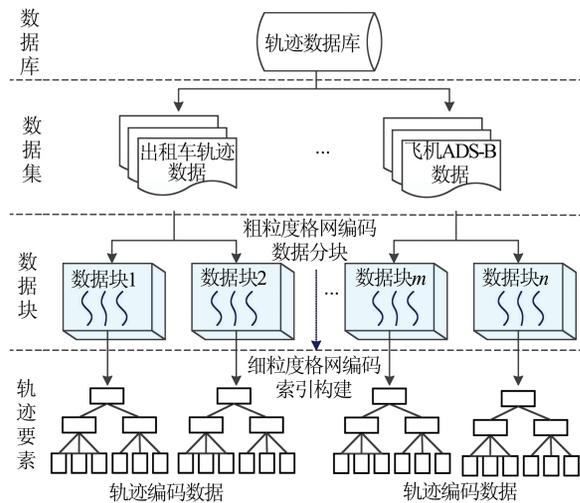


图4 轨迹编码数据的多级索引机制

Fig.4 Multi-level Indexing Mechanism of Trajectory Codes Data

### 1.3 地理格网模型与轨迹数据结合的优势

1) 存储管理高效且灵活。轨迹数据的一维地理格网编码表达降低了其存储管理和分析的难度,使得格网表达体系下的轨迹数据挖掘具有“数据和索引成一体、管理与分析相融合”的独特优势。具体来讲,轨迹数据和索引的一体化编码表达使得数据检索效率提高,避免了索引层到数据层的访问耗时,因此时空查询类算法与地理格网模型结合优势突出,包括轨迹查询、包含关系判定、范围查询、邻域分析、密度计算等。另外,基于格网编码实现的轨迹分析算子可以降低计

## 2 应用实例

本文以城市交通流可视化和轨迹数据相似性分析为应用实例,对所提出的轨迹数据管理与分析框架进行性能验证。

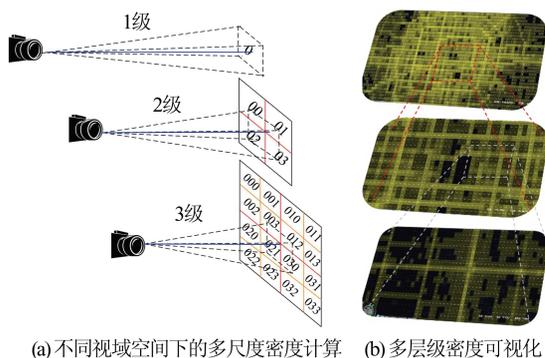
实验软件环境为 Windows10(64位)操作系统,Python运行环境,MySQL数据库,编码索引结构选择B树;硬件环境为 Dell 4-core Intel i5-4210H 2.90 GHz处理器和8 GB内存。实验数据为中国北京市约20 000辆出租车所采集的历史轨迹数据,数据量为4 656万轨迹点(约100万条轨迹),空间范围为116°00'56"E~116°43'08"E、39°22'30"N~40°04'41"N。根据轨迹数据定位精度、地理范围与道路宽度,本文基于四叉树结构将研究范围剖分至精细格网,格网尺寸约为30 m,实现轨迹数据的编码表达。

### 2.1 城市交通流多层次实时可视化

城市交通流可视化是依据传感器数据呈现

交通流在城市空间的实时分布态势,是智能交通系统的基础组成部分。大规模数据下的交通流分布态势图对于时空查询性能要求较高,而传统矢量坐标表达下的时空查询算法无法满足实时的应用需求,目前主要以静态热力图栅格切片形式展现,但其存在更新效率低、空间尺度单一等缺陷。

本文以地理格网模型为基础,将轨迹数据统一采用格网编码表达后,时空查询效率较矢量坐标下提升约一个数量级,实现了亿级数据规模下任意时空区域查询秒级响应,这使得城市交通流在任意空间尺度、任意时间段的实时动态感知成为可能。分析其内在机理在于:(1)前端视域范围的自适应动态格网表达(图 5(a))。将动态变化的视域范围自主选择合适尺度进行格网刻画,简化数据检索的查询条件;(2)后端轨迹编码数据快速查询与聚合。基于轨迹编码数据的低维度特性,并结合编码隐含的尺度信息实现快速时空聚合,生成查询网格内的交通流量。最终实现城市交通流的多层级实时可视化(图 5(b)),呈现其时空分布特征,促进智能化交通建设。



(a)不同视域空间下的多尺度密度计算 (b)多层级密度可视化

图 5 城市交通流多层级实时可视化

Fig.5 Multi-level Real-Time Visualization of Urban Traffic Flow

### 2.2 基于地理格网编码的轨迹相似性分析

轨迹相似性分析是轨迹数据挖掘的基础算法之一,决定着轨迹聚类、异常分析、伴随模式识别等分析算法的执行效率。本文基于地理格网模型改进杰卡德距离,构建了基于格网编码的相似性测度算法(similarity measurement based on grid, Grid-SIM),算法复杂度为线性,认为当两条轨迹编码重合度越高时,空间上越邻近,则两条轨迹的相似度越高。为体现 Grid-SIM 的性能,选择矢量空间下的经典相似性度量算法 DTW、EDR、LCSS、HD 作为性能对比对象,文献[42]给出了经典算法实现。

实现结果如图 6 所示,可以看出,Grid-SIM 计算效率整体远高于经典相似性度量算法,较 DTW、LCSS、EDR 平均提升约一个数量级,较 HD 提升约两个数量级,并且呈现随着查询轨迹点数量增多,查询效率提高幅度越大的趋势。这是因为经典相似性算法复杂度为平方级别,需要将两条轨迹的所有点逐一匹配计算空间距离,并选择特定的距离和作为相似性指标,因此随着轨迹点的增多时间消耗增加剧烈。而基于格网编码的 Grid-SIM 为线性复杂度,只需要判断一条轨迹内的重复编码,无需所有编码之间进行匹配,因此效率更高且较为稳定。

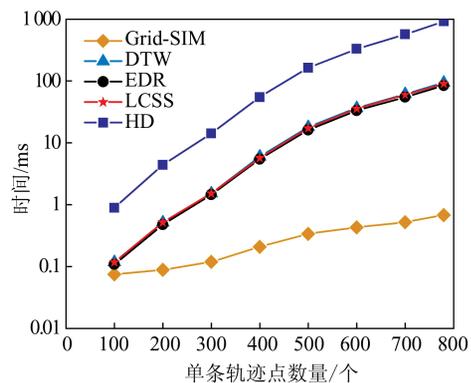


图 6 不同相似性度量方法之间在不同轨迹长度下的计算效率对比

Fig.6 Comparison of Calculation Efficiency Between Different Similarity Analysis Methods Under Different Trajectory Lengths

## 3 结语

针对现有移动对象轨迹管理与分析方法在处理大规模、复杂异构、多维动态的轨迹数据时难以满足高时效性应用需求的问题,本文提出了基于地理格网模型的轨迹数据管理与分析框架,为轨迹数据挖掘的“表达-管理-分析-应用”全链条研究提供了新的技术框架,主要包含地理格网模型、轨迹多尺度表达与组织、轨迹计算与分析、高性能计算技术及轨迹挖掘应用 5 部分。本文对各部分的实现思路和方法进行了介绍,阐述了格网模型与轨迹数据结合的优势,包括存储管理高效灵活、适合高性能计算技术和契合自动控制与智能计算需求。并以城市交通流多层级实时可视化和基于地理格网编码的相似性分析两个应用实例阐述了技术框架的应用过程,展示了该框架具有海量轨迹数据的高性能管理与分析的潜力,能够服务于城市规划、智能交通、公共卫生、社会行为等各个领域。

地理格网模型表达下的轨迹数据具有低维度、多尺度、离散型等优势,能较好地提升计算与分析效率,但对于同样的轨迹数据集,格网编码相比矢量坐标表达需要更多的存储空间,如何能在满足位置精度前提下进行数据压缩是今后需要考虑的;另外,轨迹数据本身存在粗差、缺失、冗余等问题,并且部分类型数据采集表达并非点数据,如手机信令定位单元为基站小区,这类轨迹数据的格网化编码表达需要重点考虑,以弱化数据本身误差和多样性带来的影响。未来工作中,对于综合性轨迹数据挖掘应用,必不可少需要开展轨迹数据与环境要素(道路、交通设施、感兴趣点、土地利用等)之间的协同分析,将环境要素也纳入到统一格网编码表达体系,实现基于地理格网模型的多源数据协同分析。

### 参 考 文 献

- [1] Gao Qiang, Zhang Fengli, Wang Ruijin, et al. Trajectory Big Data: A Review of Key Technologies in Data Processing[J]. *Journal of Software*, 2017, 28(4): 959-992(高强,张凤荔,王瑞锦,等. 轨迹大数据:数据处理关键技术研究综述[J]. 软件学报, 2017, 28(4): 959-992)
- [2] Zheng Y. Trajectory Data Mining: An Overview [J]. *ACM Transactions on Intelligent Systems and Technology*, 2015, 6(3): 1-41
- [3] Wu Huayi, Huang Rui, You Lan, et al. Recent Progress in Taxi Trajectory Data Mining [J]. *Acta Geodaetica et Cartographica Sinica*, 2019, 48(11): 1 341-1 356(吴华意,黄蕊,游兰,等. 出租车轨迹数据挖掘进展[J]. 测绘学报, 2019, 48(11): 1 341-1 356)
- [4] Li J, Qin Q, Xie C, et al. Integrated Use of Spatial and Semantic Relationships for Extracting Road Networks from Floating Car Data [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2012, 19: 238-247
- [5] Li J, Li Q, Zhu Y, et al. An Automatic Extraction Method of Coach Operation Information from Historical Trajectory Data [J]. *Journal of Advanced Transportation*, 2019, 2 019: 1-15
- [6] Jiang S, Fiore G A, Yang Y, et al. A Review of Urban Computing for Mobile Phone Traces: Current Methods, Challenges and Opportunities [C]//The 2nd ACM SIGKDD International Workshop on Urban Computing, New York, USA, 2013
- [7] Chen B Y, Wang Y, Wang D, et al. Understanding the Impacts of Human Mobility on Accessibility Using Massive Mobile Phone Tracking Data[J]. *Annals of the American Association of Geographers*, 2018(1): 1-19
- [8] Yang Xiping, Fang Zhixiang. Recent Progress in Studying Human Mobility and Urban Spatial Structure Based on Mobile Location Big Data[J]. *Progress in Geography*, 2018, 37(7): 880-889(杨喜平,方志祥. 移动定位大数据视角下的人群移动模式及城市空间结构研究进展[J]. 地理科学进展, 2018, 37(7): 880-889)
- [9] Li Deren, Shao Zhenfeng, Yu Wenbo, et al. Public Epidemic Prevention and Control Services Based on Big Data of Spatiotemporal Location Make Cities more Smart[J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(4): 475-487(李德仁,邵振峰,于文博,等. 基于时空位置大数据的公共疫情防控服务让城市更智慧[J]. 武汉大学学报·信息科学版, 2020, 45(4): 475-487)
- [10] Li R, Pei S, Chen B, et al. Substantial Undocumented Infection Facilitates the Rapid Dissemination of Novel Coronavirus (SARS-CoV-2) [J]. *Science*, 2020, 368(6 490): 489-493
- [11] Jia Tao, Li Qi, Ma Chu, et al. Computing the CO<sub>2</sub> Emissions of Taxi Trajectories and Exploring Their Spatiotemporal Patterns in Wuhan City[J]. *Geomatics and Information Science of Wuhan University*, 2019, 44(8): 1 115-1 123(贾涛,李琦,马楚,等. 武汉市出租车轨迹二氧化碳排放的时空模式分析[J]. 武汉大学学报·信息科学版, 2019, 44(8): 1 115-1 123)
- [12] Liu Y, Liu X, Gao S, et al. Social Sensing: A New Approach to Understanding Our Socioeconomic Environments[J]. *Annals of the Association of American Geographers*, 2015, 105(3): 512 - 530
- [13] Fang Zhixiang, Ni Yaqian, Huang Shouqian. A Multi-model Fusion Model of Individual Travel Location Prediction Using Markov and Machine Learning Methods[J]. *Geomatics and Information Science of Wuhan University*, 2020, DOI: 10.13203/j.whugis20190404(方志祥,倪雅倩,黄守倩. 融合Markov与多类机器学习模型的个体出行位置预测模型[J/OL]. 武汉大学学报·信息科学版, 2020, DOI: 10.13203/j.whugis20190404)
- [14] Zheng Y, Xie X. Learning Travel Recommendations from User-Generated GPS Traces[J]. *Transactions on Intelligent Systems and Technology*, 2011, 2(1): 1-29
- [15] Mou N, Zheng Y, Makkonen T, et al. Tourists' Digital Footprint: The Spatial Patterns of Tourist Flows in Qingdao, China [J]. *Tourism Management*, 2020, 81: 104151

- [16] Xu Jiajie, Zheng Kai, Chi Mingmin, et al. Trajectory Big Data: Data, Applications and Techniques [J]. *Journal on Communications*, 2015, 36(12): 97-105(许佳捷,郑凯,池明旻,等. 轨迹大数据:数据、应用与技术现状[J]. 通信学报,2015,36(12): 97-105)
- [17] Xiang Longgang, Gao Meng, Wang Dehao, et al. Geohash-Trees: An Adaptive Index Which can Organize Large-Scale Trajectories [J]. *Geomatics and Information Science of Wuhan University*, 2019, 44(3): 436-442(向隆刚,高萌,王德浩,等. Geohash-Trees:一种用于组织大规模轨迹的自适应索引[J]. 武汉大学学报·信息科学版,2019,44(3): 436-442)
- [18] Gong Xi, Pei Tao, Sun Jia, et al. Review of the Research Progresses in Trajectory Clustering Methods [J]. *Progress in Geography*, 2011, 30(5): 522-534(龚玺,裴韬,孙嘉,等. 时空轨迹聚类方法研究进展[J]. 地理科学进展,2011,30(5): 522-534)
- [19] Zhao Zhujun, Ji Genlin. Research Progress of Spatialtemporal Trajectory Classification [J]. *Journal of Geo-Information Science*, 2017, 19(3): 289-297(赵竹珺,吉根林. 时空轨迹分类研究进展[J]. 地球信息科学学报,2017,19(3): 289-297)
- [20] Li Y, Bailey J, Kulik L. Efficient Mining of Platoon Patterns in Trajectory Databases [J]. *Data and Knowledge Engineering*, 2015, 100: 167-187
- [21] Monreale A, Pinelli F, Trasarti R, et al. Wherenext: A Location Predictor on Trajectory Pattern Mining [C]//The 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, USA, 2009
- [22] Li Z, Ding B, Han J, et al. Mining Periodic Behaviors for Moving Objects [C]//The 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, USA, 2010
- [23] Li J, Wang J, Zhang J, et al. A Probabilistic Approach to Detect Mixed Periodic Patterns from Moving Object Data [J]. *Geoinformatica*, 2016, 20(4): 715-739
- [24] Knorr E M, Ng R T, Tucakov V. Distance-Based Outliers: Algorithms and Applications [J]. *The VLDB Journal*, 2000, 8(3/4): 237-253
- [25] Lu Feng, Liu Kang, Chen Jie. Research on Human Mobility in Big Data Era [J]. *Journal of Geo-Information Science*, 2014, 16(5): 665-672(陆锋,刘康,陈洁. 大数据时代的人类移动性研究[J]. 地球信息科学学报,2014,16(5): 665-672)
- [26] Cai Ruichu, Lin Fengji, Hao Zhifeng, et al. Research on Index Construction and Storage Method for Trajectory Stream Data [J]. *Computer Engineering*, 2020, DOI: 10.19678/j. issn. 1000-3428. 0057108(蔡瑞初,林峰极,郝志峰,等. 面向轨迹流数据的索引构建与存储方法研究[J]. 计算机工程,2020, DOI:10.19678/j. issn. 1000-3428. 0057108)
- [27] Zhang Y, Ren S, Liu Y, et al. A Framework for Big Data Driven Product Lifecycle Management [J]. *Journal of Cleaner Production*, 2017, 159(15): 229-240
- [28] Fan Q, Zhang D, Wu H, et al. A General and Parallel Platform for Mining Co-movement Patterns over Large-Scale Trajectories [J]. *Proceedings of the VLDB Endowment*, 2016, 10(4): 313-324
- [29] Ma Q, Yang B, Qian W, et al. Query Processing of Massive Trajectory Data Based on Mapreduce [C]//The 1st International Workshop on Cloud Data Management, Association for Computing Machinery, New York, USA, 2009
- [30] Xu T, Zhang X, Claramunt C, et al. TripCube: A Trip-Oriented Vehicle Trajectory Data Indexing Structure [J]. *Computers, Environment and Urban Systems*, 2018, 67: 21-28
- [31] Zhang Ping, Li Bijun, Zheng Ling, et al. A Similar Trajectory Extraction Method Based on Improved LCSS [J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(4): 550-556(张萍,李必军,郑玲,等. 一种基于改进 LCSS 的相似轨迹提取方法[J]. 武汉大学学报·信息科学版,2020,45(4): 550-556)
- [32] Liu Min, Li Mei, Xu Xiaoyu, et al. An Improved Map Matching Algorithm Based on HMM Model [J]. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 2018, 54(6): 1 235-1 241(刘旻,李梅,徐晓宇,等. 一种基于 HMM 模型改进的地图匹配算法[J]. 北京大学学报(自然科学版),2018,54(6): 1 235-1 241)
- [33] Li Deren, Shao Zhenfeng, Zhu Xinyan. Spatial Information Multi-grid and Its Typical Application [J]. *Geomatics and Information Science of Wuhan University*, 2004, 29(11): 945-950(李德仁,邵振峰,朱欣焰. 论空间信息多级格网及其典型应用[J]. 武汉大学学报·信息科学版,2004,29(11): 945-950)
- [34] Goodchild M F. Reimagining the History of GIS [J]. *Annals of GIS*, 2018, 24(1): 1-8
- [35] Zhou Chenghu, Ou Yang, Ma Ting. Progresses of Geographical Grid Systems Researches [J]. *Progress in Geography*, 2009, 28(5): 657-662(周成虎,欧阳,马廷. 地理格网模型研究进展[J]. 地理科学进展,2009,28(5): 657-662)

- [36] Chen Shupeng, Chen Qiuxiao, Zhou Chenghu. Grid Mapping and Grid Computing[J]. *Science of Surveying and Mapping*, 2002(4): 1-6(陈述彭,陈秋晓,周成虎. 网格地图与网格计算[J]. 测绘科学, 2002(4): 1-6)
- [37] Mahdavi-Amiri A, Alderson T, Samavati F. A Survey of Digital Earth[J]. *Computers and Graphics*, 2015, 53: 95-117
- [38] Zheng Y, Liu J, Li J, et al. Design of Fine Management System for Civil Aviation Airspace Resources Based on Spatiotemporal Grid Model[C]//The 1st International Conference on Civil Aviation Safety and Information Technology, Kunming, China, 2019:547-555
- [39] Zhao Xuesheng, Bai Jianjun. Hierarchical Model of Global Discrete Grids Based on Diamonds[J]. *Journal of China University of Mining and Technology*, 2007, 36(3): 397-401(赵学胜,白建军. 基于菱形块的全球离散格网层次建模[J]. 中国矿业大学学报, 2007, 36(3): 397-401)
- [40] Sun W, Cui M, Zhao X, et al. A Global Discrete Grid Modeling Method Based on the Spherical Degenerate Quadtree [C]//International Workshop on Geoscience and Remote Sensing, Shanghai, China, 2008
- [41] Cheng C, Ren F, Pu G, et al. An Introduction to Spatial Information Subdivision Organization [M]. Beijing: Science Press, 2012
- [42] Besse P C, Guillouet B, Loubes J, et al. Review and Perspective for Distance Based Trajectory Clustering[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2016, 17(11): 3 306-3 317

## Trajectory Data Management and Analysis Framework Based on Geographical Grid Model: Method and Application

LI Jun<sup>1</sup> LIU Juqing<sup>1</sup> ZHAO Xuesheng<sup>1</sup> HUANG Qian<sup>2</sup> SUN Wenbin<sup>1</sup>  
XU Zhihua<sup>1</sup> WANG Hao<sup>2</sup>

<sup>1</sup> College of Geoscience and Surveying Engineering, China University of Mining and Technology(Beijing), Beijing 100083, China

<sup>2</sup> Service Laboratory, Huawei Technology Co. Ltd, Beijing 100095, China

**Abstract: Objectives:** In recent years, trajectory data generated by various location-aware devices have been widely used in a host of fields such as urban planning, intelligent transportation, public health, and behavior analysis, and have produced huge social and economic value. However, the conventional vector representation model and the analysis algorithms based on it have high computational complexity and cannot meet the high time-efficient application requirements of large-scale trajectory data. To address the above problems, we propose a trajectory data management and analysis framework based on a geographical grid model, which provides a new technical framework for the entire chain of “representation-management-analysis-application” of trajectory data mining.**Methods:** This framework includes five parts: (1) Construction of geographical grid model. The Earth’s surface space is subdivided level by level according to some rules(e.g., quadtree, octree), and the spatiotemporal position is represented by grid coding.(2) Multi-scale coding representation and organization of trajectories. The trajectory data are mapped onto different spatial resolution grid levels according to the spatial position information and precision requirements to implement the spatial one-dimensional multi-scale coding representation instead of multi-dimensional vector coordinates, to reduce the difficulty of organization and management on massive trajectory data. (3) Calculation and analysis of trajectories. The concept of geographical grid calculation is proposed, and the existing complex algorithms are modified by using low complexity coding operations (including intrinsic index, multi-scale, set operations, etc.) to accelerate the process of trajectory data mining analysis. (4) High-performance computing technologies. We further utilize the discreteness of geographical grid model, and combine the current parallel and distributed high-performance computing frameworks to achieve distributed storage and concurrent computing on trajectory big data. (5) Applications of trajectory mining. Based on this

framework, a series of high time-efficient applications are proposed to better serve the fields of urban management, intelligent transportation, and public safety. In addition, we found several advantages of combining geographical grid model and trajectory data, including large-scale data storage and management efficiency and flexibility, suitability for high-performance computing technology, and meeting the needs of automatic control and intelligent computing. **Results:** Supported by the above theoretical methods, the reliability and effectiveness of the technical framework theory and technical methods have been verified by two applications—multi-level real-time visualizing of urban traffic flow and similarity analysis based on geographical grid model. The results show that the spatiotemporal query and similarity measure under this framework are one to two order of magnitude higher than classic algorithms under the vector coordinates. It has achieved to response within seconds and real-time visualization while processing billions of trajectory data, which proves the reliability and validity of the theories and technical methods. We explain the content of the proposed trajectory data management and analysis framework from the perspectives of framework design, theoretical analysis, advantages summary, and comparative experiments. **Conclusions:** It has been verified that the framework has the potential for high-performance management and analysis of massive trajectory data. In the future, issues such as the tradeoffs between data volume and accuracy, complex network analysis, and multi-source element association analysis need to be further explored and studied, to build a complete trajectory data management and mining technology system in a geographical grid space.

**Key words:** trajectory management; trajectory analysis; geographical grid model; moving objects; data mining

**First author:** LI Jun, PhD, associate professor, specializes in analysis and mining of geographical big data. E-mail: junli@cumtb.edu.cn

**Corresponding author:** LIU Juqing, PhD candidate. E-mail: juq.liu@ student.cumtb.edu.cn

**Foundation support:** The National Natural Science Foundation of China (41971355); Foundation for Yueqi Young Scholars of China University of Mining and Technology (Beijing).

**引文格式:** LI Jun, LIU Juqing, ZHAO Xuesheng, et al. Trajectory Data Management and Analysis Framework Based on Geographical Grid Model: Method and Application[J]. Geomatics and Information Science of Wuhan University, 2021, 46(5): 640–649. DOI: 10.13203/j.whugis20200459 (李军, 刘举庆, 赵学胜, 等. 地理格网模型支持下的轨迹数据管理与分析框架: 方法与应用[J]. 武汉大学学报·信息科学版, 2021, 46(5): 640–649. DOI: 10.13203/j.whugis20200459)