

特约稿

# 面向大数据的时空数据挖掘综述

吉根林, 赵 斌

(南京师范大学计算机科学与技术学院, 江苏 南京 210023)

[摘要] 时空数据挖掘是数据挖掘领域的前沿研究课题, 正致力于开发和应用新兴的计算技术来分析海量、高维的时空数据, 揭示时空数据中的有价值知识. 本文以时空大数据为背景, 介绍数据挖掘技术产生的背景与发展、时空数据挖掘的研究现状、研究内容、应用领域、面向大数据的时空数据挖掘系统架构以及实现技术, 为相关领域的研究者提供参考.

[关键词] 时空数据挖掘, 时空大数据, 时空模式发现, 时空聚类, 时空分类, 时空异常检测

[中图分类号] TP181 [文献标志码] A [文章编号] 1001-4616(2014)01-0001-07

## A Survey of Spatiotemporal Data Mining for Big Data

Ji Genlin, Zhao Bin

(School of Computer Science and Technology, Nanjing Normal University, Nanjing 210023, China)

**Abstract:** Spatiotemporal data mining has emerged as an active research field, focusing on the development of computing technologies for the extraction of useful information and knowledge from massive and complex spatiotemporal database. This paper mainly focuses on spatiotemporal data mining for big data, introduces the background and development of data mining, the recent theoretical and applied research progress in spatiotemporal data mining, and discusses the infrastructure and technologies of spatiotemporal data mining for big data.

**Key words:** spatiotemporal data mining, spatiotemporal big data, spatiotemporal pattern mining, spatiotemporal clustering, spatiotemporal classification, spatiotemporal outlier detection

## 1 数据挖掘技术产生的背景与发展

数据挖掘(Data Mining, DM)是指从数据集中提取人们感兴趣的知识, 这些知识是隐含的、事先未知的、潜在有用的信息, 提取的知识一般可表示为概念(Concepts)、规则(Rules)、规律(Regularities)、模式(Patterns)等形式. 1989年8月在美国底特律召开的第11届国际人工智能会议上首先出现KDD(Knowledge Discovery in Database)这个术语, 1995年学术界和工业界共同成立了ACM数据挖掘与知识发现专委会, 并组织了国际数据挖掘与知识发现大会(ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD), 后者发展成为数据挖掘领域的顶级国际会议. 至今KDD大会已连续举办19届, 论文的投稿量和参会人数呈现出逐年增加的趋势. 在KDD96国际会议上, Fayyad、Piatetsky-Shapiro和Smyth将DM和KDD定义为: KDD是指从数据库中发现知识的全部过程, Data Mining则是全部过程中的一个特定步骤, 通常两者没有明确界线, KDD和DM均指从数据中挖掘知识的过程<sup>[1]</sup>. 数据挖掘是一门交叉性学科, 涉及到人工智能、机器学习、模式识别、归纳推理、统计学、数据库、高性能计算、数据可视化等多种技术. 随着各行业对大规模数据处理和深度分析需求的快速增长, 数据挖掘已成为一个引起学术界和工业界重视、具有广泛应用需求的热门研究领域.

收稿日期: 2013-08-10.

基金项目: 国家自然科学基金(40871176).

通讯联系人: 吉根林, 博士, 教授, 博士生导师, 研究方向: 数据挖掘与云计算. E-mail: glji@njjnu.edu.cn

经过20多年的发展,数据挖掘研究取得了丰硕的成果,逐渐形成了一套自己的基础理论,主要包括规则提取、模式挖掘、分类、聚类、话题学习等.国内数据挖掘的研究和国际上领先的研究机构(卡耐基梅隆大学、康奈尔大学、斯坦福大学、伊利诺伊香槟分校、密歇根大学、微软、Google等)结合非常紧密,不少在KDD上发表的论文都是国内国外合作研究的成果.从总的趋势来看,国内和国外的研究方向仍有一些不同:在某些方面存在一定的差距,同时在某些方面也有鲜明的特色.总的来说,国外的研究更偏重于数据挖掘的理论基础和交叉学科的研究;国内的研究更偏重于实际的应用,以最新的技术解决现实的问题.国内学者在解决问题方面的研究上处于国际前沿水平,发表了大量有学术影响的论文,但在定义全新的科学问题方面还缺乏开创性的成果.在基础研究方面,国内学者和国际领先的研究团队相比还有一定的差距,这个差距正在快速缩小.例如,清华大学在今年的KDD会议就有8篇科研论文被录用,这在国际上所有的研究机构里也是十分突出的成绩.KDDCUP是每年数据挖掘算法竞赛的舞台,近年已多次看到来自国内的团队(如上海交通大学、中科院等)登上领奖台.

20世纪90年代中后期,数据挖掘领域的一些较成熟的技术,如关联规则挖掘、分类、预测与聚类等被逐渐用于时间序列数据挖掘和空间数据挖掘,以发现与时间或空间相关的有价值的模式<sup>[2-4]</sup>.近年来数据挖掘研究又有拓展,已渗透到时空数据、智能交通、生物信息、医疗卫生、金融证券、多媒体数据挖掘、文本数据挖掘、Web数据、社交网络、图数据、轨迹数据及大数据等各个领域,这些领域对数据挖掘的理论与技术提出了新的挑战,是当前数据挖掘研究的重点与难点.时空数据挖掘是数据挖掘研究的前沿领域之一,已受到国际学术界和工业界的广泛关注,被SCI/EI收录的论文数量逐年上升.国际顶级会议(如数据库领域的SIGMOD、VLDB、ICDE,数据挖掘领域的SIGKDD、ICDM)和相关领域的著名国际期刊(如《IEEE Transaction on Knowledge and Engineering》、《IEEE Transaction on Geoscience and Remote Sensing》等)每年都有很多关于时空数据挖掘研究成果的报道.当前,时空数据挖掘的研究已吸引了来自GIS、时空推理、数据挖掘、机器学习和模式识别等众多领域的学者,取得了诸多研究成果.与此同时,时空数据挖掘在许多领域得到应用,如移动电子商务(基于位置的服务)、土地利用分类及地域范围预测、全球气候变化监控(如海洋温度、厄尔尼诺现象、生物量)、犯罪易发点发现、交通协调与管理(交通中的局部失稳、道路查找)、疾病监控、水资源管理、自然灾害(如台风、森林火灾)预警、公共卫生与医疗健康等<sup>[5]</sup>.本文以时空大数据为背景,介绍时空数据挖掘的研究现状、研究内容、应用领域、面向大数据的时空数据挖掘系统架构以及实现技术,为相关领域的研究者提供参考.

## 2 时空数据挖掘的研究内容

按照挖掘的任务,时空数据挖掘主要可分为以下几类:时空模式挖掘、时空聚类、时空分类、时空异常检测等.

### 2.1 时空模式挖掘

挖掘时空数据中有价值的模式,如频繁模式、周期模式、共现模式、关联模式等,一直是时空数据挖掘研究中的一个重要课题.

#### 2.1.1 时空频繁模式

人们感兴趣的是从一个时空序列里发现频繁重复的路径,即时空频繁模式.这些被发现的频繁重复模式能协助研究人员完成关于移动对象的分析、预测等任务,进而可将其应用于经营商业、旅游业和管理城市交通等方面的决策.除了对轨迹数据之外,序列挖掘对象还包括诸如时空事件数据集等其他时空数据,在这些事件数据集中不包含对象的轨迹数据,而是由不同类型事件序列组成的集合.2001年,Tsoukatos等人<sup>[6]</sup>首次提出算法DFS\_MINE用于挖掘时空数据库中的频繁序列模式.2009年,Lee等人<sup>[7]</sup>提出一种高效的图挖掘算法,通过扫描时空数据库产生映射图和轨迹信息列表,在映射图上进行深度优先遍历以寻找频繁轨迹模式.

许多移动对象(包括交通工具、动物、移动电话用户等),在固定的时间区间内总是遵循相同或近似相同的路线,展现出一定的周期性规律.这些周期模式不仅可用于压缩移动数据,还可用来预测对象未来的移动方向.这种周期模式隐含在时空数据中需要挖掘才能发现.

### 2.1.2 时空共现模式

时空共现模式是指 2 种(或 2 种以上)对象实例在空间和时间上处于近邻. 时空共现模式已在多个领域得到了应用,如军事领域中作战计划和策略制定、生态学领域中物种和污染物跟踪、交通领域中路网规划等. 目前,时空共现模式发现研究还不多,且相关算法大都是在空间共现模式基础上通过时间扩展实现的,能有效表达时空数据不确定性和噪声的共现模式挖掘算法有待进一步研究.

### 2.1.3 时空关联模式

时空关联模式主要研究空间对象随时间发生变化的规律,即在传统关联分析的基础上加上时间和空间约束,以发现时空数据中处于一定时间间隔和空间位置的关联规则. 发现这些关联模式具有重要的应用价值,如研究战场上的战术、调查动物捕食关系等. 目前多数时空关联模式挖掘方法都是传统关联规则挖掘方法的扩展,适合时空数据的高效时空关联规则挖掘算法亟待人们去研究.

## 2.2 时空聚类

时空数据聚类是指基于空间和时间相似度把具有相似行为的时空对象划分到同一组中,使组间差别尽量大,而组内差别尽量小. 时空聚类可用于天气预测、交通拥挤预测、动物迁移分析、移动计算和异常点分析等方面. 例如,气象专家研究海岸线附近或海上飓风的共同行为,发现共同子轨迹有助于提高飓风登陆预测的准确性. 2007 年, Li 等人<sup>[8]</sup>提出一种基于密度、可发现公路网中热门路线的聚类算法,该算法主要基于共享的公共交通线上移动对象的密度进行聚类. 时空聚类研究取得的成果比较丰富,但仍存在一些问题,主要包括:1) 如何最恰当地定义对象间的距离,不同的相似度函数将决定轨迹间相似度比较的严格程度;2) 现有聚类算法未考虑时间、空间约束;3) 由于轨迹数据的复杂特征,使得“什么样的聚类能够最好地表达轨迹数据”这一问题也并未完全解决.

## 2.3 时空分类

时空分类主要是基于时空对象的特征构建分类模型来预测时空对象所属类别或对象所在具体空间位置. 对于实时物流、实时交通管理、基于位置的服务和 GPS 导航等涉及时空数据的应用而言,预测单个或一组对象未来的位置或目的地是至关重要的,它使得系统能在延误的情况下采取必要的补救措施,避免拥堵,在需要时提供有用的信息. 面向时空数据的位置和轨迹预测、密度和事件预测、结合空间的时间序列预测、结合时空信息的分类等研究具有重要应用前景. 移动对象数据库利用对象的未来位置进行最近邻搜索等预测查询;在无线网络中,对移动的预测有助于高效地分配网络资源. 目前,针对移动对象轨迹数据的分类研究还比较少.

## 2.4 时空异常检测

若某一个对象和它在空间上相邻并在一段连续时间内出现的邻居有着显著的差异,则该对象称为时空异常对象. 时空异常检测旨在从时空数据中找出严重偏离正常模式的对象. 寻找这些异常模式可以为诸多现实问题提供良好的决策支持. 譬如,在气象预测方面,预测飓风路径突然变化对于提前发出疏散指令起到至关重要的作用. 在气候方面,预测某个地区不寻常的降水行为,可以让政府和个人更好地对突如其来的洪涝灾害等极端事件做好充分的准备. 比较常见的时空异常检测方法有基于距离的方法、基于规则和模式的方法、基于密度和聚类的方法等.

时空数据异常检测在很大程度上取决于待挖掘数据的尺度和分辨率. 因此,时空数据异常检测方法还需进一步考虑空间实体间存在的度量关系(如距离关系)与非度量关系(如拓扑、方向、形状关系)等因素.

# 3 面向大数据的时空数据挖掘系统架构

## 3.1 系统架构

面向时空大数据进行数据分析与挖掘是一项复杂的工程. 按照层次结构的不同划分为 4 个部分:云计算平台、时空数据准备、数据挖掘算法和时空数据分析,如图 1 所示. 云计算是整个系统架构的底层计算平台,作为目前主流的大规模并行计算平台,它有助于时空大数据的数据管理和分析处理. 借助云计算的处理能力,可以完成时空大数据的清洗、集成、选择和加载等准备过程. 为了支持数据挖掘算法中的分析型查询操作,还需要为时空数据建立索引. 此外,时空数据挖掘的系统架构还包含经典的数据挖掘算法,借助于这些方法可以实现面向时空数据的各种应用,包括基于位置服务、时空模式发现和可视化分析等任务.

### 3.2 大数据处理平台

云计算是分布式计算、并行计算和网格计算的新发展. 它利用互联网实现随时随地、按需、便捷地访问资源池, 如计算设施、存储设备、应用程序等. 目前, 主要的云计算技术来自 Apache 基金会的 Hadoop 项目、美国加州伯克利大学 AMP 实验室的 Spark 项目和 NVIDIA 公司支持的 GPU 计算. 以下分别对这些主流的云计算技术进行介绍.

#### 3.2.1 Hadoop 项目

Hadoop<sup>[9]</sup>是由 Apache 基金会支持的分布式基础架构项目, 是 Google 公司并行编程模型 MapReduce<sup>[10]</sup>的开源版本. Hadoop 采用廉价的商用机构建数据中心平台, 具有高可靠性、高扩展性、高效性、高容错性和低成本的特点. Hadoop 主要包括分布式文件存储系统(HDFS)、任务调度及集群资源管理系统(YARN)和并行计算系统(MapReduce). 学术界已经开始研究基于 Hadoop 的时空数据管理和处理平台, 文献[11]介绍了一种基于 MapReduce 的空间数据仓库系统 Hadoop-GIS, 它通过空间分割实现多种空间数据查询, 实验表明该系统明显优于并行空间数据关系系统(Parallel SDBMS).

#### 3.2.2 Spark 项目

Spark<sup>[12]</sup>是一种基于内存计算的开源集群系统, 由美国加州伯克利大学 AMP 实验室的团队所开发. 不同于 Hadoop 依靠磁盘进行数据存储和交换, Spark 采用内存作为数据计算和处理的载体. 因而, Spark 非常适用于迭代式的数据挖掘算法. 通过比较 Spark 和 Hadoop 在常用数据挖掘算法中的计算性能, 实验发现 Spark 的运算效率比 Hadoop 提高了将近 10 倍.

#### 3.2.3 GPU 技术

GPU<sup>[13]</sup>计算是运用 GPU(图形处理器)搭配 CPU 来加速通用科学和工程应用程序. GPU 计算由 NVIDIA 公司首先提出, 并已成为一种行业标准. GPU 计算并不是单纯依靠 GPU 完成, 而是和 CPU 搭配共同完成计算任务. CPU 包含几个专为串行处理而优化的核心, 而 GPU 则由数以千计更小、更节能的核心组成, 这些核心专为提供强劲的并行性能而设计. 程序的串行部分在 CPU 上运行, 而并行部分则在 GPU 上运行.

## 4 面向大数据的时空数据挖掘应用与挑战

### 4.1 时空数据挖掘的应用

时空数据挖掘应用广泛, 如交通运输、地质灾害监测与预防、气象研究、竞技体育、犯罪分析、公共卫生与医疗及社交网络应用等, 如图 2 所示.

#### 4.1.1 交通运输

道路交通状况始终是大众出行最关心的问题. 由于城市的快速发展, 人口迅速增加, 机动车规模日益扩大, 交通拥堵给人们出行造成了许多困扰. 利用时空数据挖掘技术分析车辆的运行状况以及人流的移动规律, 可以实现交通状况的跟踪和实时预报. 例如, 文献[14]提出了一种实时预测框架, 用于预测短期的交通状况, 帮助用户避免陷入交通阻塞. 又如, 国际著名的快递公司 UPS 投入巨资研发 ORION (On-Road Integrated Optimization and Navigation)<sup>[15]</sup>系统用于快递路线优化和导航. 该系统记录了 2.5 亿个地址信息, 在综合考虑包裹信息、用户个性需求及路线情况等因素后, 为快递人员提供优化的投送路

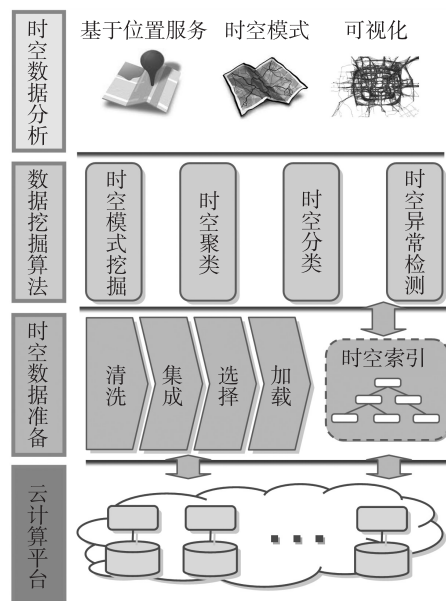


图 1 面向大数据的时空数据挖掘系统架构

Fig. 1 The architecture of spatiotemporal data mining for big data

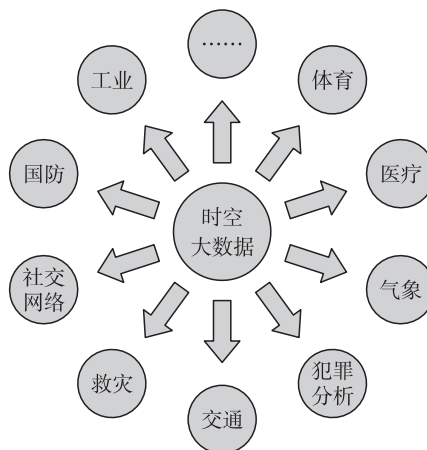


图 2 时空大数据的应用

Fig. 2 Applications of spatiotemporal data mining for big data

线. 自从 2010 年试运行至今, ORION 系统为 UPS 节省了超过 1 500 万 L 的燃料, 有效降低了 UPS 的运营成本.

#### 4.1.2 地质灾害监测与预防

地质灾害中包含了大量的时间和空间信息, 通过分析可以对将要发生的地质灾害进行监测和预防. 以地震为例, 地震是一种常见的地质灾害, 它的发生并不是完全随机的, 就地震分布而言, 在空间上大多发生在板块的边界或活动块体的边界带上, 而在时间上则经常表现为丛集性、活跃与平静的交错性等特征. 文献[16]通过对地震数据进行时空分析研究, 同时结合其他地理数据, 如应力场分布、地壳层厚度分布及温度场分布进行相关性分析, 帮助研究人员理解地震的成因和机制, 为地震预测提供依据. 为了高效处理大规模的地震数据, Meng Xiaofeng 等人研究开发了基于 GPU 的并行计算平台<sup>[17]</sup>, 实验证明可显著提高地震信号的检测效率.

#### 4.1.3 气象研究

气象对地球生态平衡和人类社会有着重要影响. 气象观测的数据既有规则的时序特征, 又有明显的空间特性, 是典型的时空数据. 时空数据分析对于气象观测与预测研究具有指导意义. 例如, 文献[18]对美国爱荷华州东部 66 个气象观测站, 从 1951 年至 2000 年的月平均最低气温数据进行时空特征分析和建模, 研究气温的时空变化趋势, 为气象的准确预报提供依据.

#### 4.1.4 竞技体育

时空数据分析技术已被引入到竞技体育中, 用以提高运动员的比赛成绩. 例如, 有研究者将时空数据分析技术用于网球比赛, 分析对手的发球轨迹, 挖掘球移动的变化规律, 总结对手的战术和策略. 这样可以帮助选手在比赛的关键时刻取得重要的比分, 从而提高比赛成绩<sup>[19]</sup>. 又如, 南加州大学的 Rajiv Maheswara 等人将数据挖掘和机器学习技术应用到 NBA 赛场, 采用特制的摄像器材以 25 次/s 的频率对球员、裁判和篮球运行进行追踪, 对球员的进攻和防守效率进行评估, 为球队的比赛和训练提供有价值的建议<sup>[20]</sup>.

#### 4.1.5 犯罪分析

犯罪是世界各国面临的日趋严重的社会问题. 通过犯罪分析研究犯罪人员行为规律, 识别犯罪模式、监测犯罪发展趋势, 可有效降低城市犯罪率. 犯罪行为在城市中的分布和发生具有一定的规律性. 犯罪地点在空间分布上的信息往往隐含了犯罪人员的住处或落脚点等信息, 而犯罪活动时间上的分布规律和犯罪类型密切相关, 例如, 文献[21]发现诈骗犯罪活动具有明显的时间选择性, 而盗窃和侵权犯罪活动在一天之内分布的比较平均. 通过对罪犯行为、动机、罪案现场、时间和被害人信息进行时空分析, 可以为犯罪案件侦查、犯罪预防提供帮助.

#### 4.1.6 公共卫生与医疗

流行病的传播规律有明显的时间和空间特征, 它可以只在某些地区发生, 也可能全球爆发. 只要条件允许, 流行病往往在适宜的环境下短时间内可能迅速发展, 广泛蔓延. 流行病一旦爆发影响广泛, 危害性大, 损失严重. 所以, 对流行病的早期检测和预警十分必要. 应用时空分析技术研究疾病数据的时空分布特征, 探讨疾病的流行区域、流行特征和流行周期, 可为公共卫生问题、卫生突发事件提供信息咨询和应对策略. 例如, 文献[22]分析全球 2009 年甲型 H1N1 流感大流行早期疫情的点空间分布模式, 通过对比历史数据, 发现国际旅行是流感传播的重要途径, 并在本次流感大流行前期主导着流感跨国传播方向. 这为以后的高致病性禽流感灾害的防控提供借鉴和依据.

#### 4.1.7 社交网络应用

近年来, 社交网络应用借助 GPS 设备记录用户轨迹数据, 通过“签到”应用分享位置信息, 如微博、人人和街旁等. 分析这些共享的位置信息可以推测用户的出行规律, 进而为用户推荐兴趣热点 (Point of Interest, POI). 例如, 文献[23]通过用户共享的 GPS 轨迹数据, 为用户推荐最感兴趣的旅游景点和游览次序.

### 4.2 时空大数据挖掘面临的挑战

#### 4.2.1 海量数据管理

为了分析处理时空大数据, 需要研究更可靠、更有效和更实用的数据管理和处理技术. 近年来, 社交网络和移动互联网的快速发展, 造成数据规模成倍扩大, 海量数据增加的速度远远超过现有的处理能力. 虽然以 MapReduce 和 Hadoop 为代表的大规模并行计算平台的出现, 为学术界提供了一条研究大数据问题

的新思路,但现有的 MapReduce 计算模型以键值对的形式组织和处理数据,并不适合处理时空数据模型。此外,Hadoop 技术无法有效支持数据挖掘中监督学习所用的迭代式计算方法,因而也无法完全满足时空数据分析的需要。另一方面,时空数据本质上是非结构化的数据,不仅包含时序数据模型,还存在图模型,例如道路网络等。基于图模型的算法时间复杂度通常比较大,对于海量数据而言,即使是  $O(N)$  的复杂度也无法承受。所以学术界和工业界必须研究面向大规模时空数据的新的数据存储管理和索引技术,以应对时空大数据的挑战。

#### 4.2.2 时空数据融合

时空数据结构复杂且来源多样,整合、清洗和转换不同来源的时空数据对于数据挖掘研究至关重要。现有的时空数据主要来源于 GPS、遥感和传感器等设备,每种设备生成的数据格式和数据形式各不相同。此外,现有的时空数据也不再局限于传统的数据形式,尤其是互联网的蓬勃发展,在文字、音频和视频等多媒体数据中同样包含了丰富的时空数据。例如,广泛覆盖城市的监控摄像头,记录了道路车辆的轨迹信息,从视频中可以还原出被监控车辆的移动轨迹。所以,对时空数据进行有效整合、清洗、转换和提取是时空数据预处理面临的重要问题。

#### 4.2.3 理论框架

与传统数据挖掘相比,时空数据挖掘研究还远未成熟,目前还缺乏一个令人满意的时空数据挖掘理论框架。构建时空数据挖掘理论框架,一方面可更好地理解时空模式需要具备的表达能力,另一方面有助于提出更有效的实现技术。理论框架应该解决以下问题:有哪些相关的模式需要挖掘;如何对这些模式进行分类;如何找到适合这些任务的挖掘算法等。

#### 4.2.4 时空推理和数据挖掘的深度结合

可度量的和不可度量的空间关系以及时间关系都需要在数据挖掘中加以考虑。然而这些时空关系往往是隐含在时空数据中的,这就需要在数据挖掘系统中结合时空推理以处理这些复杂的时空关系。

### [参考文献]

- [1] Fayyad U M, Piatetsky-Shapiro G, Smyth P. Knowledge discovery and data mining: towards a unifying framework [C]//Proceedings of KDD-96: International Conference on Knowledge Discovery and Data Mining. Portland, Oregon: AAAI Press, 1996: 82-88.
- [2] Laxman S S, Sastry P S. A survey of temporal data mining [J]. Sadhana, 2006, 31(2): 173-198.
- [3] Fu T C. A review on time series data mining [J]. Engineering Applications of Artificial Intelligence, 2011, 24(1): 164-181.
- [4] Mennis J, Guo D. Spatial data mining and geographic knowledge discovery: an introduction [J]. Computers, Environment and Urban Systems, 2009, 33(6): 403-408.
- [5] 刘大有, 陈慧灵, 齐红, 等. 时空数据挖掘研究进展 [J]. 计算机研究与发展, 2013, 50(2): 225-239.
- [6] Tsoukatos I, Gunopulos D. Efficient mining of temporal-spatial patterns [C]//Proceedings of the 7th Symp on Advances in Spatial and Temporal Databases. Berlin Heidelberg: Springer, 2001: 425-442.
- [7] Lee A J T, Chen Y A. Mining frequent trajectory patterns in temporal-spatial databases [J]. Information Sciences, 2009, 179(13): 2218-2231.
- [8] Li X, Han J, Lee J G, et al. Traffic density-based discovery of hot routes in road networks [C]//Proc of the 10th Int Conf on Advances in Spatial and Temporal Databases. Berlin Heidelberg: Springer, 2007: 441-459.
- [9] The Apache Software Foundation. Welcome to Apache Hadoop [EB/OL]. [2013-08-10] <http://hadoop.apache.org>.
- [10] Dean J, Ghemawat S. MapReduce: simplified data processing on large clusters [J]. Communications of the ACM, 2008, 51(1): 107-113.
- [11] Aji A, Wang F S, Vo H, et al. Hadoop-GIS: a high performance spatial data warehousing system over mapReduce [C]//The Proceedings of the VLDB Endowment, 2013, 6(11): 1009-1020.
- [12] Zaharia M, Chowdhury M, Franklin M J. Spark: cluster computing with working sets [C]//The 2th USENIX Workshop on Hot Topics in Cloud Computing. Berkeley, CA: USFNIX Association, 2010: 10-10.
- [13] NVIDIA 公司. 什么是 GPU 加速的计算 [EB/OL]. [2013-08-10] <http://www.nvidia.cn/object/what-is-gpu-computing-cn.html>.



- [14] Pan T L, Sumalee A, Zhong R X, et al. Short-term traffic state prediction based on temporal-spatial correlation[J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(3):1242-1254.
- [15] Konrad Alex, Forbes Staff. Meet orion, software that will save UPS millions by improving drivers's routes[EB/OL]. [2013-08-10] <http://www.forbes.com/sites/alexkonrad/2013/11/01/meet-orion-software-that-will-save-ups-millions-by-improving-drivers-routes/>.
- [16] 魏东平, 陈棋福. 地震数据时空分布的统计分析[C]//中国地球物理学会第十七届年会论文集. 北京, 2001:310-310.
- [17] Meng X F, Yu X, Peng Z G, et al. Detecting earthquakes around salton sea following the 2010 Mw7. 2 El mayor-cucapah earthquake using GPU parallel computing[C]//The International Conference on Computational Science. Nebraska, Omaha, 2012:937-946.
- [18] 刘晓晓, 孔云峰. 气象监测数据的时空特征分析与建模[J]. 地理空间信息, 2009, 7(4):104-107.
- [19] Damien Demaj. Using spatial analytics to study spatio-temporal patterns in sport[EB/OL]. [2013-08-10] <http://blogs.esri.com/esri/arcgis/2013/02/19/using-spatial-analytics-to-study-spatio-temporal-patterns-in-sport>.
- [20] Maheswaran R, Chang Y H, Henahan A, et al. Deconstructing the rebound with optical tracking data[C]//The MIT Sloan Sports Analytics Conference. Boston, 2012.
- [21] 疏学明, 颜峻, 袁宏永. 犯罪活动在一天内的发生时间规律[J]. 清华大学学报:自然科学版, 2009, 49(12):2036-2039.
- [22] 蒋之犇, 白建军, 蔡俊, 等. 2009 年甲型 H1N1 流感大流行时空分布特征分析[J]. 地球信息科学学报, 2012(6):794-799.
- [23] Zheng Y, Xie X. Learning travel recommendations from user-generated GPS traces[J]. ACM Transactions on Intelligent Systems and Technology, 2011, 2(1):1-29.

[责任编辑:严海琳]