

时空数据挖掘研究进展

刘大有 陈慧灵 齐红 杨博

(吉林大学计算机科学与技术学院 长春 130012)
(符号计算与知识工程教育部重点实验室(吉林大学) 长春 130012)
(liudayou19420601@gmail.com)

Advances in Spatiotemporal Data Mining

Liu Dayou, Chen Huiling, Qi Hong, and Yang Bo
(College of Computer Science and Technology, Jilin University, Changchun 130012)
(Key Laboratory of Symbolic Computation and Knowledge Engineering (Jilin University), Ministry of Education, Changchun 130012)

Abstract In recent years, the widespread use of the advanced technologies such as global positioning systems, sensor network and mobile devices, results in accumulation of a great amount of non-spatiotemporal data and spatiotemporal data. In addition, the processing of spatiotemporal data is more complex, which makes the increasing onerous situation of data processing tasks worse. To address these challenges, spatiotemporal data mining has emerged as an active research field, focusing on the development of theory, methodology, and practice for the extraction of useful information and knowledge from massive and complex spatiotemporal databases. Therefore, looking for effective spatiotemporal data mining methods is of great significance. This paper attempts to review the recent theoretical and applied research progress in spatiotemporal data mining and knowledge discovery. We mainly focus on spatiotemporal pattern discovery, spatiotemporal clustering, spatiotemporal anomaly detection, spatiotemporal prediction, spatiotemporal classification, and the combination of spatiotemporal data mining with reasoning. We have introduced the state-of-the-art research on spatiotemporal data mining in detail, and discussed the current major problems we are facing and its possible solutions.

Key words spatiotemporal data mining; spatiotemporal pattern mining; spatiotemporal clustering; spatiotemporal outlier detection; spatiotemporal prediction and classification

摘要 近年来,随着全球定位系统、传感器网络和移动设备等的普遍使用,非时空数据和时空数据急剧增加,加之时空数据处理更为复杂,使数据处理任务日趋繁重的形势更加严峻.因此,寻找有效的时空数据挖掘方法具有十分重要的意义.针对这一背景,主要围绕时空模式发现、时空聚类、时空异常检测、时空预测、时空分类、时空数据挖掘与推理的结合等方面,对时空数据挖掘研究的现状进行了详细介绍,对其当前所面临的一些主要问题及可能的解决方案进行了探讨.

关键词 时空数据挖掘;时空模式发现;时空聚类;时空异常检测;时空预测和分类

中图法分类号 TP181

早期的数据挖掘研究主要针对字符—数值型的商业数据.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^[1-4]或空间^[5-9]相关的有价值的模式.这些研究得到了快速发展.随着传感器网络、全球

定位系统(GPS)、手持移动设备和射频识别(RFID)等设备的普遍应用,积累了大量的移动对象数据^[1,10-13]. 此外,遥感卫星和地理信息系统(GIS)等的显著进步,使人们前所未有地获取了大量的气候数据、数字影像数据以及地理科学数据^[1,10-13]. 这些时空数据内嵌于连续空间,其样本在时间、空间上存在很强的自相关性,其中隐含的模式往往是局部的,从而使时空数据挖掘具有特殊性和复杂性. 因此,寻找有效的时空数据分析技术对于时空数据中价值时空模式的自动抽取与分析具有重要意义^[1,11].

时空数据挖掘作为一个新兴的研究领域,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^[13-14]. 近年来,时空数据挖掘已成为数据挖掘领域的研究热点,在国内外赢得了广泛关注,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 Transactions on Knowledge and Data Engineering (TKDE)》,《IEEE Transactions on Geoscience and Remote Sensing (GRSS)》,《International Journal of Geographical Information Science (IJGIS)》,《Journal of Intelligent Information Systems (JIIS)》等)每年都有很多关于时空数据挖掘研究成果的报道;同时,国际上每年都会举办有关这方面的研讨会(如数据挖掘领域权威国际会议 IEEE ICDM 从 2006 年起每年都设立“Workshop on Spatial and Spatiotemporal Data Mining”^①);此外,还有很多专题会议涉及时空

数据挖掘,例如 Symposium on Spatial and Temporal Databases (SSTD)^②, ACM International Symposium on Advances in Geographic Information Systems (ACM GIS)^③等.

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,时空数据挖掘也在许多领域得到应用,如移动电子商务(基于位置的服务)^[15-16]、土地利用分类及地域范围预测^[17-19]、全球气候变化监控(如海洋温度、厄尔尼诺现象、生物量)^[20-22]、犯罪易发点发现^[23]、交通协调与管理(交通中的局部失稳、道路查找)^[24]、疾病监控^[25]、水资源管理^[26-27]、自然灾害(如台风、森林火灾)预警^[28-29]、公共卫生与医疗健康^[30-33]等.

本文对时空数据挖掘领域的研究现状以及目前面临的主要问题进行了分析,试图为该研究方向勾画出一个较为全面和清晰的轮廓,为数据挖掘、机器学习和 GIS 等相关领域的研究者提供有益的参考.

1 时空数据挖掘

按照数据挖掘的定义^[12],可将时空数据挖掘定义为从具有海量、高维、高噪声和非线性等特性的时空数据中提取出隐含的、人们事先不知道的、但又潜在有用的信息及知识的过程. 时间维和空间维为数据挖掘任务增加了额外的复杂性,使得时空数据挖掘研究更具挑战性. 作为一个新兴的研究领域,经过近年的快速发展,时空数据挖掘研究领域出现了

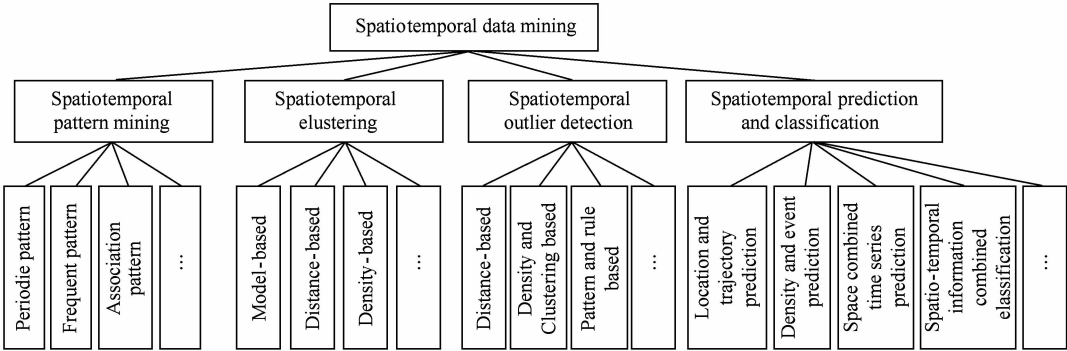


Fig. 1 The classification chart of spatiotemporal data mining tasks.

图 1 时空数据挖掘任务分类图

① <http://i.cs.hku.hk/~nikos/sstdm06/>, <http://spatial.nvc.cs.vt.edu/sstdm07>, <http://www.ornl.gov/sci/knowledgediscovery/sstdm-08/>, <http://www.ornl.gov/sci/knowledgediscovery/sstdm-09/>
② <http://www.informatik.uni-trier.de/~ley/db/conf/ssd/index.html>
③ <http://www.sigspatial.org/sigspatial-conferences>

很多有价值的工作^[34-35].按照挖掘的任务主要可分为以下几类:时空模式发现、时空聚类、时空异常检测、时空预测和分类等,如图1所示.除这几类较常见的方法外,时空推理与数据挖掘结合的方法也逐渐受到人们的关注.本文将主要围绕这些方面来阐述时空数据挖掘领域的最新进展.

1.1 时空模式发现

发现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.

1.1.1 时空频繁模式(spatiotemporal frequent pattern)

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决策.

2001年,Tsoukatos等人^[36]首次提出在时空数据库中挖掘频繁序列模式,并提出一种新算法DFS_MINE.该算法利用深度优先搜索,通过扫描数据库来挖掘序列模式,能同时支持在不同空间粒度上进行序列挖掘.2005年,Cao等人^[37]提出一种时空频繁序列模式挖掘方法,首先将原始序列转换成一系列子序列片段,然后用子串树结构和改进的Apriori算法挖掘频繁区域.该方法主要特点在于,按形状相似程度和空间上的远近程度对分段物体进行重组,同时利用空间实体间的邻近性质提高时空序列模式的搜索效率.但上述工作并未考虑轨迹序列区域间的时间信息.为了同时捕获序列区域间的时间和空间信息,2007年,Giannotti等人^[38]在原有带时间标注的序列上^[39]进一步考虑空间维信息,引入一种新的轨迹模式(T模式),用于描述同时包含相同时间段和空间位置的连续轨迹序列的频繁行为,进而提出一种T模式挖掘算法,该算法能动态发现感兴趣的空间区域.2009年,Lee等人^[40]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.该方法在效率上比基于Apriori和PrefixSpan的频繁模式挖掘方法提高了一个数量级.但不能挖掘那些由非相连点形成的模式.

上述工作都是针对移动对象轨迹进行频繁序列模式挖掘的.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并不包含对象的轨迹数据,而是一个

由不同类型事件序列组成的集合.2006年和2008年,Huang等人^[41-42]提出2种从时空事件数据集中挖掘时空序列模式的算法STS-Miner和Slicing-STs-Miner,并提出将一个序列索引作为时空序列模式的重要性度量,利用空间统计学建立起统计解释.但这两种算法都是以固定的时间区间和空间位置假设为前提的.

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^[43-44].这种周期模式隐含在时空数据中需要挖掘才能发现.这些周期模式不仅可用于压缩移动数据^[43-44],而且还可用来预测对象未来的移动方向^[45].

2004年,Mamoulis等人^[43]将周期模式发现问题定义为从有效的非冗余频繁模式中搜索周期模式的过程.同时,针对该问题他们提出一种从历史时空数据中自顶向下检索最大周期模式的高效算法(STPMine2).此外,他们还针对所发现的周期模式,设计了一种能有效管理时空数据的新索引模式.2007年,Cao等人^[44]提出一种检索最大周期模式的高效算法,进而通过研究某个连续子区间内频繁发生的周期模式发现和轨迹被转移和扭曲的周期模式发现问题,验证了算法的实用性.但该算法只支持细时间粒度上的挖掘.2009年,Karli等人^[46]提出两种能在不同时间粒度上挖掘单个移动对象运动轨迹的周期模式发现算法.该算法首先利用基于密度的聚类方法抽取“重要区域”,然后利用一种紧凑的离散表示方法对这些重要区域采用精确和近似两种匹配,以分别支持关于粗时间粒度和细时间粒度的周期模式挖掘.2010年,Li等人^[47]提出了一种二阶段算法Periodica,对移动对象轨迹中复杂的多重交叉周期行为进行挖掘.该算法的主要思想是,先用傅里叶转换和自相关技术探测移动对象的周期,然后用层次聚类算法来统计移动对象的周期行为.该工作克服了文献^[43-44]存在的输出模式集合中存在微小偏差模式的问题.但Periodica算法在动态探测空间位置参考点时只考虑了空间信息,若同时考虑时间信息则有可能进一步提高探测的准确率.

虽然时空频繁模式挖掘取得了一些研究成果,但仍有如下一些问题未得到很好的解决:1)缺乏指导时空频繁模式挖掘算法参数设置的方法;2)尚未见考虑移动对象空间位置不确定性,或针对特殊的时空数据(如稀疏的时空数据、抽样变化率不稳定的时空数据等)的挖掘算法.

1.1.2 时空共现模式(spatiotemporal co-location pattern)

时空共现模式系指 2 种(或 2 种以上)对象实例在空间和时间上处于邻近^[48]. 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^[48]等.

目前与时空共现模式相关的大多数研究结果都是在空间共现模式^[49]基础上进行时间维扩展得到的. 2005 年, Wang 等人^[50]提出一种高效的时空共现模式发现算法 TopologyMiner, 采用深度优先搜索方式来发现共现模式, 该方法的特点在于可以在挖掘过程中消除大量的候选集. 2005 年, Yang 等人^[51]将数据的特征建模为几何对象, 并定义了多种距离度量, 将时间信息引入空间共现模式, 进而提出一种时空演化共现模式挖掘算法. 2006 年和 2008 年, Celik 等人^[48, 52]定义了一种混合时空共现模式度量方法, 提出一种混合时空共生模式算法, 并对该算法进行了正确性和完备性分析. 利用相同的方法, Celik 等人^[53]定义了一种可持续出现的时空共现模式, 并提出另一种基于距离的算法. 2006 年, Cao 等人^[54]阐明了从时空数据中挖掘不同移动对象间频繁共现子序列的问题. 该问题可揭示不同类型移动对象间相互作用的规律. 进而提出一种针对该问题的高效二阶段挖掘算法. 该算法的核心思想是先用 Hash 技术将原始轨迹转换成具有相近特征的子序列, 然后利用 Apriori 算法挖掘其中的频繁片段. 上述研究在模式挖掘过程中时间维度和空间维度是分开进行考虑的. 2008 年, Zhang 等人^[55]从结合角度考虑了时间维和空间维的度量问题, 定义了一种新的度量方法, 进而提出一种高效的时空共现模式发现算法 COSTCOP⁺.

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.

1.1.3 时空关联模式(spatiotemporal association pattern)

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了时间和空间约束, 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.

2004 年, Tao 等人^[56]提出一种结合时空索引和简图技术的方法, 该方法可大大减小搜索空间, 加快查询速度, 并可获得很高的精度. 但该方法只对时空关联规则作了简单定义, 并没有考虑空间和时间语义(如区域面积或区域间的时间间隔等)信息. 2006 年和 2008 年, Verhein 等人^[57-58]考虑了区域面积和区域间时间间隔等两方面, 完整定义了一种描述物体随时间跨区域移动的时空关联规则, 并定义了一种能高效处理不同大小区域问题的空间支持度和在空间区域中的时间模式(区域中的源、汇、固定区域和通道等模式), 并提出一种高效挖掘这些模式的算法 STAR-Miner. 为了增强关联模式的效率和可解释性, Verhein 等人^[59-60]还引入序列时空关联模式(k -STAR).

除此之外, 还有许多从不同角度挖掘时空关联模式的研究. 2005 年, Gidófalvi 等人^[61]提出一种旋转方法(pivoting)将时空关联规则挖掘任务转换为传统的购物篮分析任务, 进而用传统关联规则挖掘算法进行挖掘. 2006 年, Lee 等人^[62]通过扩展 Apriori 算法提出一种高效时空关联规则挖掘算法. 该算法能发现时空数据库中具有多层次粒度的时空关联规则. 2007 年, Yang 等人^[63]提出一种时空关联框架挖掘蛋白质折叠轨迹, 探测关键的折叠活动和公共路径. 该框架用联络图结合三维结构信息进行辅助分析, 在动态折叠过程中, 首先在联络图中挖掘时空关联模式, 然后研究这些模式如何随着折叠轨迹进行演变. 2008 年, Yang 等人^[64]提出一种高效的能捕获折叠过程演变内在规律、识别折叠路径中的三维结构单元的算法, 进而提出一种解析结构单元间演变过程和相互作用的时空关联框架. 2009 年, Leong 等人^[23]提出一种动态模式分析框架. 该框架可探测 3 种动态模式: 不同时间点的相同空间模式、两个区域间的相互作用模式, 与特定事件、地点以及时间点相关的频繁关联规则. 2009 年, Hamed 等人^[65]提出一种在线时空关联规则挖掘框架. 该框架将传统关联规则扩展到多元连续数据, 并利用时空相关性对观测变量进行智能推理.

除以上这些比较常见的时空模式外, 从时空数据中还可以挖掘出许多其他模式, 如群体移动模式(motion pattern)^[66-68]、流动模式(flow pattern)^[69]以及迁移模式(migration motif)^[70]等.

总之, 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.

1.2 时空聚类(spatiotemporal clustering)

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。

1999年,Gaffney等人^[71]提出一种基于回归混合模型的概率算法对轨迹数据进行聚类,并利用期望最大化算法确定模型中簇的隶属度。基于此,2003年,Chudova等人^[72]将对象的时间和空间轨迹的漂移也作为混合模型的参数来研究数据的聚类。2003年,Alon等人^[73]提出一种可表达对象在两个相邻位置转换的马尔可夫模型对簇进行表示。以上方法都是基于模型来聚类轨迹数据,目标是希望获得一种能描述数据的全局模式。其中一些方法依赖于定义的多元密度分布,并寻找模型的拟合参数。和基于模型的方法不同,基于距离的方法主要定义一种基于距离的相似度函数来对轨迹进行聚类。2006年,Nanni等人^[74]提出基于时态语义和时间聚焦方法的轨迹聚类算法。该算法的核心思想是用一种基于密度的聚类算法对轨迹进行聚类,其中距离采用给定时间段轨迹间空间距离的平均值,并将时间间隔作为距离的参数。基于密度的聚类方法主要在于定义一个密度阈值来区分相关数据项和噪音。2007年,Birant等人^[75]在核心对象、噪声对象和相邻簇等方面对DBSCAN算法进行扩展,提出一种新的基于密度的时空聚类算法ST-DBSCAN。该算法可根据对象的非空间、空间和时间值来发现簇。2007年,Li等人^[76]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。除上述方法外,研究人员还提出了许多基于移动微簇探测移动轨迹数据中移动簇的方法。2004年,Li等人^[77]提出能捕捉移动对象时空规律的微聚类算法,使移动微簇在空间上保持微小的距离,同时可检测移动微簇之间的碰撞等重要事件。2005年,Kalnis等人^[78]将移动对象看作是关联时间信息的空间位置集合,并将聚类问题转化为发现具有共同时间片的空间簇的问题。但微聚类方法也存在一定的局限性,譬如,微簇的定义限制了算法只能找到球形簇,在簇与簇发生重叠期间算法不易将簇分开。此外,如果移动对象的速度频

繁变化,更新分离和合并操作将会占据整个算法大部分时间。

上述大部分方法都将整条轨迹视为最小的簇,其主要缺点是不能捕获不同轨迹中具有相似功能的子轨迹片段。2007年,Lee等人^[79]提出一种可以发现轨迹中公共子轨迹的划分与聚合(partition-and-group)框架TRACCLUS,其核心是基于最小描述长度原则的轨迹划分算法。该算法首先将轨迹表示成不包含明确时间信息的点序列,使用启发式算法将每个轨迹划分成准线性片段的集合,然后利用基于密度的聚类方法对相似片段进行聚类。为了能够适应海量数据的聚类,增量方法被引入以增加聚类效率。2007年,Elnekave等人^[80]使用一种紧凑表示方法表示移动轨迹,并定义了一种新的相似性度量方法,提出一种增量聚类算法来发现相似移动对象的演化簇。2010年,Li等人^[81]提出针对移动轨迹数据的增量聚类框架。该框架主要包含在线微观聚类维护和离线宏观聚类创建两部分,实验结果表明该框架可以同时达到很高的效率和聚类精度。2008年,Palma等人^[82]提出一种基于速度的聚类算法CB-SMoT,该方法考虑轨迹的语义属性(背景地理信息),能从单条轨迹中挖掘出重要位置。该方法的主要特点在于对时空数据的语义信息进行了考虑;此外,与从轨迹集合中发现簇的常规方法相比,该方法主要从单条轨迹中发现一些具有重要意义的位置。2009年,Pelekis等人^[83]提出一种考虑不确定因素的轨迹数据聚类方法。该方法共包含3个步骤:首先提出一种直观的模糊向量轨迹表示法;其次提出一种基于距离度量法,能发现运动群中心轨迹的算法CenTra;最后提出一种改进的模糊C均值聚类算法对轨迹进行聚类。该方法的主要特点在于对时空数据内在不确定性以及数据中存在的大量噪音进行了考虑。2010年,Rosswog等人^[84]针对噪声环境中的移动对象,提出一种基于动态密度的时空聚类算法DDBC,引入关系图表示物体的关系历史,以增量地估计移动对象间关系的强度,并提出改进的DBSCAN算法,在关系图中挖掘强相关的对象簇。

尽管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。

1.3 时空异常检测(spatiotemporal outlier detection)

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的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的原因对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入其来的洪涝灾害等极端事件做好充分的准备.

2004 年, Adam 等人^[85]提出一种基于距离的异常检测算法, 将相似度定义为物体间的空间关系及其他语义关系的度量. 该算法可检测高速公路交通资源监控和水资源监控过程中的异常情况. 2006 年, Jin 等人^[86]同时考虑时间和空间, 利用马氏距离对交通异常事件进行实时检测. 该方法具有很强的自适应学习能力和快速的响应能力. 2009 年, Das 等人^[87]利用基于距离和邻居的异常探测方法从全球气候系统中检测出时空异常模式, 这些异常模式不仅可以很好概括 2004 至 2006 年来全球气候的突变情况, 而且很好地解释了诸如地球上某个具体区域发生干旱和严重降雨等极端事件. 同时实验中还发现地球上某几个区域间存在遥相关模式(比如亚热带地区的海洋表面温度导致南极洲半岛的气候异常), 结果表明这些异常和当年的新闻报道完全一致. 以上工作基本都是基于距离的时空异常检测方法. 除此之外, 结合基于密度和聚类的方法在时空异常检测中也比较常见. 2006 年, Birant 等人^[88]提出一种适合处理大数据集的时空异常检测算法. 2008 年, Lee 等人^[89]提出一种基于分区的异常轨迹检测框架, 将轨迹分成线段的集合, 对轨迹中异常点的外围线段进行检测; 基于该框架提出一种轨迹异常检测算法. 该算法主要由双层轨迹分割策略和基于距离和密度的混合方法组成. 2006 年, Cheng 等人^[90]从不同的时间和空间尺度探测异常情况, 提出一种包含分类、聚集、比较和验证 4 个步骤的多尺度方法. 该方法充分考虑了地理现象的语义和动态属性, 通过评估连续空间和时间两个尺度间的变换来探测沿海地貌的异常情况. 近年来出现了一些新的时空异常检测方法, 譬如, 2008 年, Li 等人^[91]提出一种基于规则和模式的异常检测框架. 该框架主要采用离散的片段模式表示对象的轨迹, 将抽取的相关特征形成层次特征空间; 然后设计基于规则的分类器, 进而在多层次粒度上对有效规则进行学习. 2009

年, Bu 等人^[92]提出一种监视连续轨迹流异常情况的新框架. 该框架首先利用轨迹的局部连续特点来建立局部簇, 然后通过高效的减枝策略对异常情况进行监视, 进而提出一种分段度量索引结构来重新安排局部簇的添加顺序, 进一步减少了时间代价. 2010 年, Gaber 等人^[93]提出一种近似算法 Approx-Grid Top- k . 该算法通过逼近离群区域的差异性可快速发现各个时间段内的离群点; 在该算法基础上他们又提出一种新算法 Outstretch, 并应用该新算法在南美降水数据中探测到很多随位置、形状和大小变化的时空异常模式.

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.

1.4 时空预测和分类(spatiotemporal prediction and classification)

时空预测和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.

1.4.1 位置和轨迹预测

位置预测研究主要来自两个领域: 数据库系统和无线网络. 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.

2000 年, Saltenis 等人^[94]提出一种新的基于 R^* 树的索引结构—TPR 树, 可在 1 维、2 维和 3 维空间内对移动对象进行索引, 从而支持对移动对象当前和未来位置的高效查询. 2002 年, Tao 等人^[95]在传统的空间查询(窗口查询、 k -近邻和空间连接查询)上引入时间参数, 根据查询类型定义距离函数, 提出基于时间参数的最近邻空间查询. 2003 年, Tao 等人^[96]提出 TPR 树的改进版本 TPR * 树, 通过一组改进的构造算法使其能更好地适应动态对象的特性, 实验表明改进的 TPR * 树显著提高了查询性能. 2006 年, Benetis 等人^[97]提出平面连续运动点的反向最近

邻和最近邻查询算法,并给出当数据库更新(插入或删除一条记录)时查询的增量式更新方法.2007年,张伟等人^[98]提出一种基于概率模型的时空区域查询处理方法,将移动对象未来可能出现的位置定义为一个随机变量,并能计算移动对象在两种不同的运动模式下满足查询的概率值.上述研究都假定对象是在一个无障碍的空间中进行线性运动,如船、飞机、天气现象等.然而,却不适用于街道网络,其中的对象会在很短的时间间隔内改变方向和速度.

对于街道网络、无线通信网络等不稳定的环境,更需要对不久的将来对象的移动进行预测.1997年,Biesterfeld等人^[99]通过训练神经网络来预测移动网络中对象的位置.2003年,Liang等人^[100]提出基于对象位置和速率的高斯-马尔可夫模型,用于预测无线PCS(个人通信系统)中用户的移动模式.2003年,Katsaros等人^[101]使用聚类算法挖掘PCS用户的移动模式,以预测用户的移动并动态分配资源.2005年,Liou等人^[102]提出基于用户位置、速度、加速度和方向的移动模型和一个能预测移动主机(x, y)坐标的神经网络预测器.2005年,Yavas等人^[103]应用序列模式挖掘方法从历史轨迹数据中发现移动规则,并用其预测用户的移动.2006年,Song等人^[104]对无线网络中的位置预测方法进行了全面的研究和比较.2009年,Monreale等人^[105]提出WhereNext方法.该方法使用预先抽取的轨迹模式,从某个区域的轨迹模式中学习决策树,然后通过树中查找最匹配的路径来预测新轨迹的下一个位置.

除了位置预测之外,另一个重要的任务是预测移动对象最有可能的路线和目的地.例如,如果知道用户将路过的位置,基于位置的服务可以提供更有针对性的服务.2003年,Karimi等人^[106]将转移矩阵应用于个人偏好数据,预测某个人在给定时段内最有可能的路线和目的地.2005年,Laasonen等人^[107]在Karimi等人提出的模型中结合驻留时间,将用户花费相对较多时间的地方作为所有可能的目的地;同时对历史路线进行聚类并与当前轨迹进行比较,在最相似的轨迹中预测目的地.

1.4.2 密度、事件和结合空间的时间序列预测

某个区域的对象密度定义为在给定时间点该区域内对象数与该区域大小之比.这是一些对象随时间变化而呈现出的一个全局特征.密度预测能给交通领域的应用带来很多好处,例如,交通管理系统若能识别出交通中的密集区域,就可采取措施及时缓

解这些瓶颈.

2003年,Hadjieleftheriou等人^[108]在计算密度时定义了时空立方体的概念,每个单元包含给定区域(x 和 y 轴)某个时间间隔(z 轴)的密度;并假定对象是线性移动的,通过推断来计算未来的密度.2004年,Sun等人^[109]考虑了每个空间区域的变化,将之前区域密度的加权和作为后续区域的密度;权值遵循指数平滑过程,强调最近区域密度的影响,减少时间较远的区域密度的权重.

2001年,Brown等人^[110]研究了时空事件预测问题,定义了一个密度转换模型.该模型可根据历史数据,结合密度估计(发现重要特征和时空地点)来预测给定时间和空间的概率密度,譬如基于过去事件的地点、时间和社会经济等特征预测给定区域和时间段内犯罪发生的概率.

结合空间的时间序列预测是从时间的角度来考虑时空数据.与传统时间序列不同的是,与空间有关的时空序列彼此不是独立的,而是空间相关的.

2003年,Pokrajac等人^[111]提出一种根据从统一网格上采样的短期历史数据预测时空数据的方法,基于时空自回归模型,使用每个采样点空间邻居的最近历史响应值来预测随后时间层的响应值.同年Li等人^[112]提出时空结合预测框架STIFF.该框架首先构造时间序列模型以获取每个空间独立区域的时间特性;然后构造神经网络模型发现隐藏的空间相关性;最后基于统计回归结合时间和空间预测获得综合预测.同年Zhang等人^[113]使用空间自相关性,即通过对地点分组过滤掉大批相关性低于给定阈值的元素对,来降低对结合空间的时间序列数据库进行相关性分析时的计算复杂性.

1.4.3 结合时空信息的分类

时空数据分类可获得非常有价值的信息.例如,区分观光者和本地工作者的不同路径,可使基于位置的服务能为用户提供更适合的信息.目前,针对移动对象轨迹数据的分类研究还比较少,相关研究还包括时间序列分类.

1998年,Keogh等人^[114]提出时间序列的分段线性表示方法,并根据每段的重要性赋予相应的权重,可用于分类、聚类和关联性反馈等方面.2001年,Geurts等人^[115]研究了时间序列分类问题,通过搜寻和组合时间序列中的模式进行分类,并提出分类模式搜寻和分类规则组合构建方法.2007年,Bashir等人^[116]结合高斯混合模型和隐马尔可夫模型,提出一种使用移动轨迹识别对象活动的分类算法.

2008年, Lee等人^[117]提出基于层次特征的轨迹分类方法, 首先基于区域进行聚类, 再进行轨迹聚类, 这种结合区域簇和轨迹簇的分类方法可获得较高的分类精度. 2009年, Schultz等人^[118]提出基于模糊推理系统、遗传算法和遗传规划的时空分类方法, 通过对一系列卫星地图的分析来预测给定时间某地的地形地貌.

目前, 时空预测和分类研究面临的主要问题包括: 1) 不稳定、多变环境下(如交通应用中的城市网络)的预测问题要求算法很快适应结构变化, 获得新的移动模式; 2) 结合不同来源、不同质量的数据(如无线网络中基于蜂窝的移动数据, GPS的精确定位数据)的研究还较少见.

1.5 时空推理和数据挖掘的结合

时空数据结构复杂、关系繁多且难以描述. 目前, 多数数据挖掘系统容纳的时空关系过于简单, 使其能发现的时空知识十分有限. 时空数据的许多定性推理规则(如时间关系方面, 先与后等传递属性; 空间关系方面, 远与近等度量属性), 以及时空本体都提供了独立的领域知识来源, 应成为时空模式挖掘时需要考虑的因素. 2008年, 国际权威学者在《Machine Learning》上撰文指出推理和学习的深度结合是未来10年机器学习和数据挖掘面临的挑战之一^[119]. 目前, 时空推理和数据挖掘的结合研究甚少^[14, 120].

2000年, Clementini等人^[121]针对空间数据挖掘方法不支持不确定性的问题, 提出宽边界概念, 用来描述空间对象间的拓扑关系及其不确定性, 并提出基于宽边界的多层空间关联规则挖掘方法. 2004年, 刘大有、王生生等人^[122]以RCC理论为基础, 结合模糊逻辑, 提出面向空间数据库的近似区域空间关系模型, 基于该模型提出多层空间关联规则挖掘算法QSR SAR, 其表达能力、精度和效率均优于宽边界方法. 同年, Santos等人^[123]实现的地理数据库知识发现系统PADRAO, 根据地理数据库中存储的相邻区域间的空间关系, 使用定性时空推理可推导出数据挖掘阶段所需的非相邻区域间的所有关系. 2005年, Yang等人^[124-125]提出一种时空模式挖掘和推理相结合的框架, 定义了描述空间实体关系的时空对象关联模式SOAP, 并给出了挖掘算法, 基于SOAP模式提出多种推理策略. 2006年, 王生生、刘大有等人^[126]提出结合多方面(拓扑、方向、距离和尺寸)空间推理的空间数据挖掘系统SRSDM, 并成功应用于农业土壤数据库. 2007年, Mukhlash等

人^[127]在数据预处理阶段根据用户指定的空间关系(如拓扑关系和距离关系)将空间数据转化为特殊形式的数据进行存储, 然后使用传统的关联规则挖掘算法Apriori和FP-Growth从中挖掘空间关联规则.

上述研究工作中, 时空推理和数据挖掘的结合往往是串行的, 因此存在如下问题: 1) 现有方法首先应用时空推理抽取出所有的时空关系, 再进行时空知识挖掘, 而事实上所需的关系可能只是所有时空关系的一个子集, 这将导致大量时间和资源的浪费. 一个好的解决方案是将推理集成到挖掘过程中, 在需要生成时空关系或进行模型评估时再应用时空推理. 2) 在实际应用中, 有时耗费大量时间和资源学习出一个表达能力很强的模型, 但在随后的推理中却不得使用近似推理, 这将造成精度的大幅下降. 因此, 挖掘过程中需要对模型表达能力和推理能力进行权衡折中.

2 面临的挑战

尽管时空数据挖掘研究在近几年引起了人们的广泛关注并得到快速发展, 但与传统数据挖掘相比, 时空数据挖掘研究还远未成熟, 仍有诸多问题需要进一步探索, 主要表现在以下几个方面:

1) 理论框架. 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. 清晰定义的理论框架将会给该研究领域带来理论上的指导.

2) 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. 时空推理和数据挖掘的深度结合将会使所发现的时空模式包含更多的时空信息, 增强时空模式的可理解性; 同时还可显著提高挖掘的效率和质量. 但应注意的是: 抽取数据中隐含定义的时空关系必然会引入某种程度的不确定性和模糊性, 这将对数据挖掘的结果带来很大影响; 同时需要在模型表达能力和推理能力之间进行适当折中.

3) 多粒度时空挖掘. 空间粒度和时间粒度对所发现模式的强度有直接影响. 时空数据挖掘方法应该能在不同粒度水平发现最感兴趣的模式(例如最支持的、最可信的). 目前, 很少有数据挖掘技术能自动支持这种多级分析. 用户通常在不同表示粒度下重复独立的实验, 且粗粒度水平得到的结果并没有被用来控制细粒度水平的搜索(反之亦然).

4) 结合背景知识. 空间和时间数据的许多定性推理规则(如时间关系 after 和 before 的传递性)和时空本体(特别是关于地理空间的本体)提供了诸多有价值的知识来源. 时空对象间的关系可表示为一种自然的依赖关系(例如港口临近水体). 如果这些依赖关系被用于约束新模式的搜索, 时空数据挖掘算法的可扩展性将得到大幅提高. 如何表达这些关系, 如何在时空数据挖掘系统中结合背景知识仍然是一个开放性问题.

5) 不确定性处理. 不确定性是时空数据的固有特征. 由于物理和技术方面的局限, 不确定性往往产生于数据采集和存储过程中. 尽管可假定时间能被精确地进行表示, 但随着所采用技术的不同, 空间位置的不确定性却可能在几米(GPS)到几千米(GSM)之间变化. 另外, 采样率对不确定性也有很大影响. 为了能在一个给定的水平维持空间不确定性, 当对象在空间移动得越快时则关于其空间位置的频度就应越高. 这些因素都需在时空数据挖掘和融合过程中予以充分考虑.

6) 元规则挖掘. 很多时候无法直接从原始数据中挖掘时空模式. 例如, 由于某种原因(如出于隐私和安全的考虑)原始数据无法获得, 或者原始数据只能保留很短的时间(如流数据), 或者数据量巨大导致挖掘算法复杂性急剧增加. 一个好的解决办法是应用元规则挖掘(或称元学习), 即在先前挖掘获得的模式或模型中发现知识^[128]. 元规则可获得更接近人类直觉的知识. 如何将最早在时间数据挖掘领域提出的元规则挖掘扩展到时空数据挖掘领域将是一个有价值的研究方向.

7) 高性能时空挖掘算法. 随着应用领域的不断延伸和数据规模的急剧增长, 迫切需要研究高性能时空模式发现方法. 例如适合于非结构化数据和 Web 数据的时空模式挖掘算法; 带约束条件的轨迹聚类方法; 移动对象轨迹的增量聚类、子轨迹聚类算法; 结合多源数据的高精度协作预测方法; 结构多变环境下移动模式发现方法等. 与此同时, 还需结合新的应用领域, 探索新的时空模式及其高效挖掘算法.

譬如, 将时空数据挖掘应用于生物信息学领域, 可协助分析蛋白质中的折叠轨迹和检测折叠路径, 对基因调控网络的进化行为进行建模和预测等.

3 总 结

本文主要围绕时空模式发现、时空聚类、时空异常检测、时空预测、时空分类、以及时空数据挖掘和推理深度结合等方面综述了时空数据挖掘领域近年来的主要研究进展. 重点讨论了各种方法取得的最新研究成果以及面临的一些主要问题, 并给出了一些可能采取的研究方案. 除此之外, 其他与时空数据挖掘相关的问题还包括时空模式和现象的可视化^[129-130]、查询^[131-133], 时空数据的表示^[134-135], 时空聚集计算^[56, 136-137]以及时空数据索引^[138-141]等. 这些问题的研究也会对时空数据挖掘方法的效率和应用产生较大的影响, 因此在这些问题上也应给予足够的关注.

参 考 文 献

- [1] Antunes C M, Oliveira A L. Temporal data mining: An overview [C] // Proc of the Int Workshop on Temporal Data Mining (KDD2001). New York: ACM, 2001: 1-13
- [2] Roddick J F, Spiliopoulou M. A survey of temporal knowledge discovery paradigms and methods [J]. IEEE Trans on Knowledge and Data Engineering, 2002, 14(4): 750-767
- [3] Laxman S, Sastry P S. A survey of temporal data mining [J]. Sadhana, 2006, 31(2): 173-198
- [4] Fu T C. A review on time series data mining [J]. Engineering Applications of Artificial Intelligence, 2011, 24(1): 164-181
- [5] Koperski K, Adhikary J, Han J. Knowledge discovery in spatial databases: Progress and challenges. [C] // Proc of the ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery. New York: ACM, 1996: 55-70
- [6] Shekhar S, Zhang P, Huang Y, et al. Data Mining: Next Generation Challenges and Future Directions [M]. Cambridge, MA: MIT Press, 2004: 357-380
- [7] Shekhar S, Zhang P, Huang Y. Data Mining and Knowledge Discovery Handbook [M]. Berlin: Springer, 2010: 837-854
- [8] Miller H J, Han J. Geographic Data Mining and Knowledge Discovery [M]. London: Taylor and Francis, 2001
- [9] Mennis J, Guo D. Spatial data mining and geographic knowledge discovery—An introduction [J]. Computers, Environment and Urban Systems, 2009, 33(6): 403-408

- [10] Miller H J, Han J. Geographic Data Mining and Knowledge Discovery [M]. Boca Raton: CRC Press, 2009
- [11] Han J, Russ B A, Kumar V, et al. Emerging scientific applications in data mining [J]. ACM Communication, 2002, 45(8): 54-58
- [12] Han J, Kamber M. Data mining: Concepts and Techniques [M]. San Francisco: Morgan Kaufmann, 2000
- [13] Han J. Research frontiers in advanced data mining technologies and applications, in advances in knowledge discovery and data mining [G] //LNCS 4426; Proc of the 11th Pacific-Asia Conf on Knowledge Discovery and Data Mining. Berlin: Springer, 2007: 1-5
- [14] Andrienko G, Malerba D, May M, et al. Mining spatio-temporal data [J]. Journal of Intelligent Information Systems, 2006, 27(3): 187-190
- [15] Lee W H, Tseng S S, Wang C H, et al. Discovering the radio signal coverage hole and weak coverage area in mobile network by spatiotemporal data mining on location-based services [J]. Fundamenta Informaticae, 2010, 98(1): 33-47
- [16] Gidófalvi G, Pedersen T. Mining long, sharable patterns in trajectories of moving objects [J]. GeoInformatica, 2009, 13(1): 27-55
- [17] Hong X, Liu Y, Liu D, et al. Design and implementation of multi-source data mining system for land use [C] //Proc of the Int Symp on Spatial Analysis, Spatial-Temporal Data Modeling, and Data Mining. Bellingham, WA: SPIE, 2009: 74921J-1-74921J-7
- [18] Goo J, Ranga Raju V, Joydeep G. Spatially adaptive classification and active learning of multispectral data with gaussian processes [C] //Proc of the 2009 IEEE Int Conf on Data Mining Workshops. Piscataway, NJ: IEEE, 2009: 597-603
- [19] Tang J, Wang L, Yao Z. Spatio-temporal urban landscape change analysis using the Markov chain model and a modified genetic algorithm [J]. Int Journal of Remote Sensing, 2007, 28(15): 3255-3271
- [20] Huang Y P, Kao L J, Sandnes F E. Efficient mining of salinity and temperature association rules from ARGO data [J]. Expert Systems with Applications, 2008, 35(1): 59-68
- [21] Vatsavai R R. BioMon: A Google Earth based continuous biomass monitoring system [C] //Proc of the 17th ACM SIGSPATIAL Int Conf on Advances in Geographic Information Systems. New York: ACM, 2009: 536-537
- [22] Ganguly A R, Steinhäuser K. Data mining for climate change and impacts [C] //Proc of the 2008 IEEE Int Conf on Data Mining Workshops. Piscataway, NJ: IEEE, 2008: 385-394
- [23] Leong K, Li J, Chan S, Ng V. An application of the dynamic pattern analysis framework to the analysis of spatial-temporal crime relationships [J]. Journal of Universal Computer Science, 2009, 15(9): 1852-1870
- [24] Lee W H, Tseng S S, Tsai S H. A knowledge based real-time travel time prediction system for urban network [J]. Expert Systems with Applications, 2009, 36(3): 4239-4247
- [25] Zhang M, Kong X, Wallstrom G. Simulation of multivariate spatial-temporal outbreak data for detection algorithm evaluation [G] //LNCS 5354; Proc of the Int Workshop on Biosurveillance and Biosecurity (BioSecure 2008). Berlin: Springer, 2008: 155-163
- [26] Su S, Zhi J, Lou L, et al. Spatio-temporal patterns and source apportionment of pollution in Qiantang River(China) using neural-based modeling and multivariate statistical techniques [J]. Physics and Chemistry of the Earth, Parts A/B/C, 2011, 36(9): 379-386
- [27] Hsu K C, Li S T. Clustering spatial-temporal precipitation data using wavelet transform and self-organizing map neural network [J]. Advances in Water Resources, 2009, 33(2): 190-200
- [28] Kitamoto A. Spatio-temporal data mining for typhoon image collection [J]. Journal of Intelligent Information Systems, 2002, 19(1): 25-41
- [29] Cheng T, Wang J. Integrated spatio-temporal data mining for forest fire prediction [J]. Transactions in GIS, 2008, 12(5): 591-611
- [30] Rainham D, McDowell I, Krewski D, et al. Conceptualizing the healthscape: Contributions of time geography, location technologies and spatial ecology to place and health research [J]. Social Science & Medicine, 2010, 70(5): 668-676
- [31] Cervone G, Stefanidis A, Franzese P, et al. Spatiotemporal modeling and monitoring of atmospheric hazardous emissions using sensor networks [C] //Proc of the IEEE Int Conf on Data Mining Workshops (ICDMW'09). Piscataway, NJ: IEEE, 2009: 571-576
- [32] McNally R J Q, Parker L. Environmental factors and childhood acute leukemias and Lymphomas [J]. Leukemia & Lymphoma, 2006, 47(4): 583-598
- [33] Meliker J R, Sloan C D. Spatio-temporal epidemiology: Principles and opportunities [J]. Spatial and Spatio-temporal Epidemiology, 2011, 2(1): 1-9
- [34] Roddick J, Hornsby K, Spiliopoulou M. An updated bibliography of temporal, spatial, and spatio-temporal data mining research [G] //LNCS 2007; Proc of the 1st Int Workshop on Temporal, Spatial, and Spatio-Temporal Data Mining(TSDM 2000). Berlin: Springer, 2001: 147-163
- [35] Nanni M, Kuijpers B, Korner C, et al. Mobility, Data Mining and Privacy [M]. Berlin: Springer, 2008: 267-296
- [36] Tsoukatos I, Gunopulos D. Efficient mining of spatiotemporal patterns [G] //LNCS 2121; Proc of the 7th Int Symp on Advances in Spatial and Temporal Databases (SSTD 2001). Berlin: Springer, 2001: 425-442
- [37] Cao H, Mamoulis N, Cheung D W. Mining frequent spatio-temporal sequential patterns [C] //Proc of the 5th IEEE Int Conf on Data Mining. Piscataway, NJ: IEEE, 2005: 82-89

- [38] Giannotti F, Nanni M, Pinelli F, et al. Trajectory pattern mining [C] //Proc of the 13th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2007: 330-339
- [39] Giannotti F, Nanni M, Pedreschi D. Efficient mining of sequences with temporal annotations [C] //Proc of the 6th SIAM Int Conf on Data Mining (SDM'06). Philadelphia, PA: SIAM, 2006: 346-357
- [40] Lee A J T, Chen Y-A, Ip W-C. Mining frequent trajectory patterns in spatial-temporal databases [J]. Information Sciences, 2009, 179(13): 2218-2231
- [41] Huang Y, Zhang L, Zhang P. Finding sequential patterns from a massive number of spatio-temporal events [C] //Proc of the 6th SIAM Int Conf on Data Mining (SDM'06). Philadelphia, PA: SIAM, 2006: 633-637
- [42] Huang Y, Zhang L, Zhang P. A framework for mining sequential patterns from spatio-temporal event data sets [J]. IEEE Trans on Knowledge and Data Engineering, 2008, 20(4): 433-448
- [43] Mamoulis N, Cao H, Kollios G, et al. Mining, indexing, and querying historical spatiotemporal data [C] //Proc of the 10th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2004: 236-245
- [44] Cao H, Mamoulis N, Cheung D W. Discovery of periodic patterns in spatiotemporal sequences [J]. IEEE Trans on Knowledge and Data Engineering, 2007, 19(4): 453-467
- [45] Hoyoung J, Qing L, Heng Tao S, et al. A hybrid prediction model for moving objects [C] //Proc of the 24th Int Conf on Data Engineering. Piscataway, NJ: IEEE, 2008: 70-79
- [46] Karli S, Saygin Y. Mining periodic patterns in spatio-temporal sequences at different time granularities [J]. Intelligent Data Analysis, 2009, 13(2): 301-335
- [47] Li Z, Ding B, Han J, et al. Mining periodic behaviors for moving objects [C] //Proc of the 16th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2010: 1099-1108
- [48] Celik M, Shekhar S, Rogers J P, et al. Mixed-drove spatiotemporal co-occurrence pattern mining [J]. IEEE Trans on Knowledge and Data Engineering, 2008, 20(10): 1322-1335
- [49] Shekhar S, Huang Y. Discovering spatial co-location patterns: A summary of results [G] //LNCS 3882: Proc of the 7th Int Symp on Advances in Spatial and Temporal Databases(SSTD 2001). Berlin: Springer, 2001: 236-256
- [50] Wang J, Hsu W, Lee M L. A framework for mining topological patterns in spatio-temporal databases [C] //Proc of the 14th ACM Int Conf on Information and Knowledge Management. New York: ACM, 2005: 429-436
- [51] Yang H, Parthasarathy S, Mehta S. A generalized framework for mining spatio-temporal patterns in scientific data [C] //Proc of the 11th ACM SIGKDD Int Conf on Knowledge Discovery in Data Mining. New York: ACM, 2005: 716-721
- [52] Celik M, Shekhar S, Rogers J P, et al. Mixed-drove spatio-temporal co-occurrence pattern mining: A summary of results [C] //Proc of the 6th Int Conf on Data Mining(ICDM'06). Piscataway, NJ: IEEE, 2006: 119-128
- [53] Celik M, Shekhar S, James P R, et al. Sustained emerging spatio-temporal co-occurrence pattern mining: A summary of results [C] //Proc of the 18th IEEE Int Conf on Tools with Artificial Intelligence (ICTAI'06). Piscataway, NJ: IEEE, 2006: 106-115
- [54] Cao H, Mamoulis N, Cheung D W. Discovery of collocation episodes in spatiotemporal data [C] //Proc of the 6th Int Conf on Data Mining(ICDM'06). Piscataway, NJ: IEEE, 2006: 823-827
- [55] Zhang Z, Wu W. Composite spatio-temporal co-occurrence pattern mining [G] //LNCS 5258: Proc of the 3rd Int Conf on Wireless Algorithms, Systems, and Applications(WASA 2008). Berlin: Springer, 2008: 454-465
- [56] Tao Y, Kollios G, Considine J, et al. Spatio-temporal aggregation using sketches [C] //Proc of the 20th Int Conf on Data Engineering. Piscataway, NJ: IEEE, 2004: 214-225
- [57] Verhein F, Chawla S. Mining spatio-temporal association rules, sources, sinks, stationary regions and thoroughfares in object mobility databases [G] //LNCS 3882: Proc of the 11th Int Conf on Database Systems for Advanced Applications (DASFAA 2006). Berlin: Springer, 2006: 187-201
- [58] Verhein F, Chawla S. Mining spatio-temporal patterns in object mobility databases [J]. Data Mining and Knowledge Discovery, 2008, 16(1): 5-38
- [59] Verhein F. k -STARs: Sequences of spatio-temporal association rules [C] //Proc of the 6th IEEE Int Conf on Data Mining Workshops. Piscataway, NJ: IEEE, 2006: 387-394
- [60] Verhein F. Mining complex spatio-temporal sequence patterns [C] //Proc of SIAM Int Conf on Data Mining(SDM 2009). Philadelphia, PA: SIAM, 2009: 605-616
- [61] Gidófalvi G, Pedersen T. Spatio-temporal rule mining: Issues and techniques [G] //LNCS 3589: Proc of the 7th Int Conf on Data Warehousing and Knowledge Discovery (DaWaK 2005). Berlin: Springer, 2005: 275-284
- [62] Lee E M H, Chan K C C. Discovering association patterns in large spatio-temporal databases [C] //Proc of the 6th IEEE Int Conf on Data Mining Workshops. Piscataway, NJ: IEEE, 2006: 349-354
- [63] Yang H, Parthasarathy S, Ucar D. A spatio-temporal mining approach towards summarizing and analyzing protein folding trajectories [J]. Algorithms for Molecular Biology, 2007, 2(3): 1-16

- [64] Yang H, Han L. An effective approach for identifying evolving three-dimensional structural motifs in protein folding data [G] //LNCS 5139; Proc of the 4th Int Conf on Advanced Data Mining and Applications (ADMA 2008). Berlin: Springer, 2008; 342-354
- [65] Hamed C, Le G. An online spatio-temporal association rule mining framework for analyzing and estimating sensor data [C] //Proc of the 2009 Int Database Engineering & Applications Symp. New York: ACM, 2009; 217-226
- [66] Hwang S Y, Liu Y H, Chiu J K, et al. Mining mobile group patterns: A trajectory-based approach [G] //LNCS 3518; Proc of the 9th Pacific-Asia Conf on Advances in Knowledge Discovery and Data Mining (PAKDD 2005). Berlin: Springer, 2005; 145-146
- [67] Gudmundsson J, Kreveld M V, Speckmann B. Efficient detection of motion patterns in spatio-temporal data sets [C] //Proc of the 12th Annual ACM Int Workshop on Geographic Information Systems. New York: ACM, 2004; 250-257
- [68] Hyunjin Y, Shahabi C. Accurate discovery of valid convoys from moving object trajectories [C] //Proc of the IEEE Int Conf on Data Mining Workshops(ICDMW'09). Piscataway, NJ: IEEE, 2009; 636-643
- [69] Wang J, Hsu W, Lee M L. FlowMiner: Finding flow patterns in spatio-temporal databases [C] //Proc of the 16th IEEE Int Conf on Tools with Artificial Intelligence. Piscataway, NJ: IEEE, 2004; 14-21
- [70] Du X, Jin R, Ding L, et al. Migration motif: A spatial-temporal pattern mining approach for financial markets [C] //Proc of the 15th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2009; 1135-1144
- [71] Graffney S, Smyth P. Trajectory clustering with mixtures of regression models [C] //Proc of the 5th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 1999; 63-72
- [72] Chudova D, Gaffney S, Mjolsness E, et al. Translation-invariant mixture models for curve clustering [C] //Proc of the 9th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2003; 79-88
- [73] Alon J, Sclaroff S, Kollios G, et al. Discovering clusters in motion time-series data [C] //Proc of the 2003 IEEE Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2003; 375-381
- [74] Nanni M, Pedreschi D. Time-focused clustering of trajectories of moving objects [J]. Journal of Intelligent Information Systems, 2006, 27(3): 267-289
- [75] Birant D, Kut A. ST-DBSCAN: An algorithm for clustering spatial-temporal data [J]. Data & Knowledge Engineering, 2007, 60(1): 208-221
- [76] 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: Springer, 2007; 441-459
- [77] Li Y, Han J, Yang J. Clustering moving objects [C] //Proc of the 10th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2004; 617-622
- [78] Kalnis P, Mamoulis N, Bakiras S. On discovering moving clusters in spatio-temporal data [C] //Proc of the 9th Int Conf on Advances in Spatial and Temporal Databases. Berlin: Springer, 2005; 923-923
- [79] Lee J G, Han J, Whang K Y. Trajectory clustering: A partition-and-group framework [C] //Proc of the 2007 ACM SIGMOD Int Conf on Management of Data. New York: ACM, 2007; 593-604
- [80] Elnekave S, Last M, Maimon O. Incremental clustering of mobile objects [C] //Proc of the 23rd Int Conf on Data Engineering Workshop. Piscataway, NJ: IEEE, 2007; 585-592
- [81] Li Z, Lee J G, Li X, et al. Incremental clustering for trajectories [C] //Proc of the 15th Int Conf on Database Systems for Advanced Applications Database Systems for Advanced Applications. Berlin: Springer, 2010; 32-46
- [82] Palma A T, Bogorny V, Kuijpers B, et al. A clustering-based approach for discovering interesting places in trajectories [C] //Proc of the 2008 ACM Symp on Applied Computing. New York: ACM, 2008; 863-868
- [83] Pelekis N, Kopanakis I, Kotsifakos E E, et al. Clustering trajectories of moving objects in an uncertain world [C] //Proc of the 9th Int Conf on Data Mining. Piscataway, NJ: IEEE, 2009; 417-427
- [84] Rosswog J, Ghose K. Efficiently detecting clusters of mobile objects in the presence of dense noise [C] //Proc of the 2010 ACM Symp on Applied Computing. New York: ACM, 2010; 1095-1102
- [85] Adam N R, Janeja V P, Atluri V. Neighborhood based detection of anomalies in high dimensional spatio-temporal sensor datasets [C] //Proc of the 2004 ACM Symp on Applied Computing. New York: ACM, 2004; 576-583
- [86] Jin Y, Dai J, Lu C. Spatial-temporal data mining in traffic incident detection [C] //Proc of SIAM Int Conf on Data Mining Workshop on Spatial Data Mining (SDM 2006). Philadelphia, PA: SIAM, 2006; 20-22
- [87] Das M, Parthasarathy S. Anomaly detection and spatio-temporal analysis of global climate system [C] //Proc of the 3rd Int Workshop on Knowledge Discovery from Sensor Data. New York: ACM, 2009; 142-150
- [88] Kut A, Birant D. Spatio-temporal outlier detection in large databases [J]. Journal of Computing and Information Technology, 2006, 14(4): 214-224

- [89] Lee J G, Han Jiawei, Lee Xiaolei. Trajectory outlier detection: A partition-and-detect framework [C] //Proc of the 24th Int Conf on Data Engineering. Piscataway, NJ: IEEE, 2008: 140-149
- [90] Cheng T, Li Z. A multiscale approach for spatio-temporal outlier detection [J]. Transactions in GIS, 2006, 10(2): 253-263
- [91] Li X, Han J, Kim S, et al. Anomaly detection in moving object [G] //Studies in Computational Intelligence 135: Intelligence and Security Informatics, Techniques and Applications. Berlin: Springer, 2006: 187-201
- [92] Bu Y, Chen L, Fu A W C, et al. Efficient anomaly monitoring over moving object trajectory streams [C] //Proc of the 15th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2009: 159-168
- [93] Gaber M, Vatsavai R, Omataomu O, et al. Spatio-temporal outlier detection in precipitation data [G] //LNCS 5840: Proc of the 2nd Int Workshop on Knowledge Discovery from Sensor Data(Sensor-KDD 2008). Berlin: Springer, 2008: 115-133
- [94] Saltenis S, Jensen C S, Leutenegger S T, et al. Indexing the positions of continuously moving objects [C] //Proc of the 2000 ACM SIGMOD Int Conf on Management of data. New York: ACM, 2000: 331-342
- [95] Tao Y, Papadias D. Time-parameterized queries in spatio-temporal databases [C] //Proc of the 2002 ACM SIGMOD Int Conf on Management of data. New York: ACM, 2002: 334-345
- [96] Tao Y, Papadias D, Sun J. The TPR* -tree: An optimized spatio-temporal access method for predictive queries [C] //Proc of the 29th Int Conf on Very Large Data Bases. San Francisco: Morgan Kaufmann, 2003, 790-801
- [97] Benetis R, Jensen C S, Karciauskas G, et al. Nearest and reverse nearest neighbor queries for moving objects [J]. The VLDB Journal, 2006, 15(3): 229-249
- [98] Zhang Wei, Li Jianzhong, Liu Yu. A probabilistic model based predictive spatio-temporal range query processing [J]. Journal of Software, 2007, 18(2): 279-290 (in Chinese)
(张炜, 李建中, 刘禹. 一种基于概率模型的预测性时空区域查询处理[J]. 软件学报, 2007, 18(2): 279-290)
- [99] Biesterfeld J, Ennigrou E, Jobmann K. Location prediction in mobile networks with neural networks [C] //Proc of the Int Workshop on Applications of Neural Networks to Telecommunications (IWANN'T'97). Melbourne, Australia: Lawrence Erlbaum Associates, 1997: 207-214
- [100] Liang B, Haas Z J. Predictive distance-based mobility management for multidimensional PCS networks [J]. IEEE/ACM Trans on Networking, 2003, 11(5): 718-732
- [101] Katsaros D, Nanopoulos A, Karakaya M, et al. Clustering mobile trajectories for resource allocation [C] //Proc of the 5th Int Symp on Intelligent Data Analysis (IDA 2003). Berlin: Springer, 2003: 319-329
- [102] Liou S C, Huang Y M. Trajectory predictions in mobile networks [J]. International Journal of Information Technology, 2005, 11(11): 109-122
- [103] Yavas G, Katsaros D, Ulusoy O, et al. A data mining approach for location prediction in mobile environments [J]. Data & Knowledge Engineering, 2005, 54(2): 121-146
- [104] Song Libo, Kotz D, Jain R, et al. Evaluating next-cell predictors with extensive Wi-Fi mobility data [J]. IEEE Trans on Mobile Computing, 2006, 5(12): 1633-1649
- [105] Monreale A, Pinelli F, Trasarti R, et al. WhereNext: A location predictor on trajectory pattern mining [C] //Proc of the 15th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining(KDD'09). New York: ACM, 2009: 637-645
- [106] Karimi H A, Liu X. A predictive location model for location-based services [C] //Proc of the 11th ACM Int Symp on Advances in Geographic Information Systems. New York: ACM, 2003: 126-133
- [107] Laasonen K. Clustering and prediction of mobile user routes from cellular data [G] //LNCS 2637: Proc of the 9th European Conf on Principles and Practice of Knowledge Discovery in Databases. Berlin: Springer, 2005: 569-576
- [108] Hadjieleftheriou M, Kollios G, Gunopulos D, et al. On-line discovery of dense areas in spatio-temporal databases [G] //LNCS 2750: Proc of the 8th Int Symp of Advances in Spatial and Temporal Databases (SSTD 2003). Berlin: Springer, 2003: 306-324
- [109] Sun J, Papadias D, Tao Y, et al. Querying about the past, the present, and the future in spatio-temporal databases [C] //Proc of the 20th Int Conf on Data Engineering (ICDE'04). Piscataway, NJ: IEEE, 2004: 202-213
- [110] Brown D E, Liu H, Xue Y. Mining preferences from spatial-temporal data [C] //Proc of the 1st SIAM Int Conf on Data Mining(SDM'01). Philadelphia, PA: SIAM, 2001: 1-17
- [111] Pokrajac D, Hoskinson R L, Obradovic Z. Modeling spatial-temporal data with a short observation history [J]. Knowledge and Information Systems, 2003, 5(3): 368-386
- [112] Li Z, Dunham M, Xiao Y. STIFF: A forecasting framework for spatiotemporal data mining multimedia and complex data [G] //LNCS 2797: Proc of the PAKDD Workshop on Mining Multimedia and Complex Data (KDMCD 2002). Berlin: Springer, 2003: 183-198
- [113] Zhang P, Huang Y, Shekhar S, et al. Correlation analysis of spatial time series datasets: A filter-and-refine approach [G] //LNCS 2637: Proc of the 11th Pacific-Asia Conf on Knowledge Discovery and Data Mining. Berlin: Springer, 2003: 563-563

- [114] Keogh E, Pazzani M. An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback [C] //Proc of the 4th Int Conf on Knowledge Discovery and Data Mining(KDD'98). New York: ACM, 1998: 239-241
- [115] Geurts P. Pattern extraction for time series classification principles of data mining and knowledge discovery [G] // LNCS 2168: Proc of the 5th European Conf on Principles of Data Mining and Knowledge Discovery. Berlin: Springer, 2001: 115-127
- [116] Bashir F I, Khokhar A A, Schonfeld D. Object trajectory-based activity classification and recognition using hidden Markov models [J]. IEEE Trans on Image Processing, 2007, 16(7): 1912-1919
- [117] Lee J G, Han J, Li X, et al. TraClass: Trajectory classification using hierarchical region-based and trajectory-based clustering [J]. Proc VLDB Endow, 2008, 1(1): 1081-1094
- [118] Schultz R E O, Centeno T M, Selleron G, et al. A soft computing-based approach to spatio-temporal prediction [J]. Int Journal of Approximate Reasoning, 2009, 50(1): 3-20
- [119] Dietterich T G, Domingos P, Getoor L, et al. Structured machine learning: The next ten years [J]. Machine Learning, 2008, 73(1): 3-23
- [120] Ceci M, Appice A. Spatial associative classification: Propositional vs structural approach [J]. Journal of Intelligent Information Systems, 2006, 27(3): 191-213
- [121] Clementini E, Di Felice P, Koperski K. Mining multiple-level spatial association rules for objects with a broad boundary [J]. Data & Knowledge Engineering, 2000, 34(3): 251-270
- [122] Liu Dayou, Wang Shengsheng, Yu Qiangyuan, et al. A multi-level spatial association rules mining algorithm based on qualitative spatial reasoning [J]. Journal of Computer Research and Development, 2004, 41(4): 565-570 (in Chinese)
(刘大有, 王生生, 虞强源, 等. 基于定性空间推理的多层空间关联规则挖掘算法[J]. 计算机研究与发展, 2004, 41(4): 565-570)
- [123] Santos M Y, Amaral L A. Mining geo-referenced data with qualitative spatial reasoning strategies [J]. Computers & Graphics, 2004, 28(3): 371-379
- [124] Yang H, Parthasarathy S, Mehta S. Mining spatial object associations for scientific data [C] //Proc of the 19th Int Joint Conf on Artificial Intelligence (IJCAI 2005). San Francisco: Morgan Kaufman, 2005: 902-907
- [125] Yang H, Parthasarathy S, Mehta S. Towards association based spatio-temporal reasoning [C] //Proc of the 19th IJCAI Workshop on Spatiotemporal Reasoning. San Francisco: Morgan Kaufman, 2005: 77-85
- [126] Wang S S, Liu D Y, Wang X Y, et al. Spatial reasoning based spatial data mining for precision agriculture [G] // LNCS 3842: Proc of Int Workshops on Advanced Web and Network Technologies, and Applications (APWeb 2006). Berlin: Springer, 2006: 506-510
- [127] Mukhlash I, Sitohang B. Spatial data preprocessing for mining spatial association rule with conventional association mining algorithms [C] //Proc of the Int Conf on Electrical Engineering and Informatics (ICEEI 2007). Bandung: ICEEI, 2007: 531-534
- [128] Roddick J F, Spiliopoulou M, Lister D, et al. Higher order mining [J]. ACM SIGKDD Explorations Newsletter, 2008, 10(1): 5-17
- [129] Andrienko N, Andrienko G, Gatalsky P. Exploratory spatio-temporal visualization: An analytical review [J]. Journal of Visual Languages & Computing, 2003, 14(6): 503-541
- [130] Compieta P, Di Martino S, Bertolotto M, et al. Exploratory spatio-temporal data mining and visualization [J]. Journal of Visual Languages & Computing, 2007, 18(3): 255-279
- [131] Ortale R, Ritacco E, Pelekis N, et al. The DAEDALUS framework: Progressive querying and mining of movement data [C] //Proc of the 16th ACM SIGSPATIAL Int Conf on Advances in Geographic Information Systems. New York: ACM, 2008: 1-4
- [132] Manco G, Baglioni M, Giannotti F, et al. Mobility, Data Mining, and Privacy: Geographic Knowledge Discovery [M]. Berlin: Springer, 2008
- [133] Pfoser D, Jensen C S, Theodoridis Y. Novel approaches in query processing for moving object trajectories [C] //Proc of the 26th Int Conf on Very Large Data Bases. San Francisco: Morgan Kaufmann, 2000: 395-406
- [134] Güting R, Bohlen M, Erwig M, et al. A foundation for representing and querying moving objects [J]. ACM Trans on Database Systems(TODS), 2000, 25(1): 1-42
- [135] Erwig M, Schneider M. Spatio-temporal predicates [J]. IEEE Trans on Knowledge and Data Engineering, 2002, 14(4): 881-901
- [136] Lopez I, Snodgrass R, Moon B. Spatiotemporal aggregate computation: A survey [J]. IEEE Trans on Knowledge and Data Engineering, 2005, 17(2): 271-286
- [137] Timko I, Bohlen M H, Gamper J. Sequenced spatio-temporal aggregation in road networks [C] //Proc of the 12th Int Conf on Extending Database Technology: Advances in Database Technology. New York: ACM, 2009: 48-59
- [138] Kollios G, Gunopulos D, Tsotras V J. On indexing mobile objects [C] //Proc of the 18th ACM SIGMOD-SIGACT-SIGART Symp on Principles of Database Systems. New York: ACM, 1999: 261-272

[139] Altenis S, Jensen C S, Leutenegger S T, et al. Indexing the positions of continuously moving objects [C] //Proc of the 2000 ACM SIGMOD Int Conf on Management of data. New York: ACM, 2000: 331-342

[140] Hadjieleftheriou M, Kollios G, Tsotras V J, et al. Efficient indexing of spatiotemporal objects [C] //Proc of the 8th Int Conf on Extending Database Technology: Advances in Database Technology. Berlin: Springer, 2002: 251-268

[141] Tao Y, Papadias D, Sun J. The TPR* -tree: An optimized spatio-temporal access method for predictive queries [C] // Proc of the 29th Int Conf on Very Large Data Bases. San Francisco: Morgan Kaufmann, 2003: 790-801



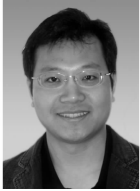
Liu Dayou, born in 1942. Professor and PhD supervisor. Senior member of China Computer Federation. His main research interests include knowledge engineering and expert system, multi-agent system, spatial-temporal reasoning, data mining and Bayesian network (liudy@jlu.edu.cn).



Chen Huiling, born in 1983. PhD. His main research interests center on data mining and machine learning (chenhuiling.jlu@gmail.com).



Qi Hong, born in 1970. PhD and associate professor. His main research interest is data mining (qihong@jlu.edu.cn).



Yang Bo, born in 1974. Professor and PhD supervisor. His main research interests include agent system, data mining and complex network analysis (ybo@jlu.edu.cn).