

Yidong Duixiang Guji Shuju Wajue Jishu
China University of Mining and Technology Press



移动对象轨迹数据挖掘技术

袁冠 张艳梅 著



中国矿业大学出版社

移动对象轨迹数据挖掘技术

袁冠 张艳梅 著



中国矿业大学出版社

内 容 简 介

本书较为系统地介绍了移动对象数据挖掘的现状、技术,从微观、宏观、模式发现、原型系统开发四个角度介绍了移动对象数据挖掘的基本原理和方法,并结合当前位置服务的需求,介绍了主流的移动对象数据挖掘框架。对应移动对象数据挖掘的四个方面,本书分别介绍了典型的技术方法,帮助读者揭开移动对象数据挖掘技术的面纱。

本书可作为高校高年级本科生和研究生的选修教材,也可作为从事移动对象数据挖掘相关领域的科技人员的参考书。

图书在版编目(CIP)数据

移动对象轨迹数据挖掘技术 / 袁冠, 张艳梅著. —

徐州: 中国矿业大学出版社, 2016.11

ISBN 978 - 7 - 5646 - 3360 - 8

I. ①移… II. ①袁… ②张… III. ①数据采集

IV. ①TP274

中国版本图书馆 CIP 数据核字(2016)第 302120 号

- 书 名 移动对象轨迹数据挖掘技术
著 者 袁 冠 张艳梅
责任编辑 姜 华
出版发行 中国矿业大学出版社有限责任公司
(江苏省徐州市解放南路 邮编 221008)
营销热线 (0516)83885307 83884995
出版服务 (0516)83885767 83884920
网 址 <http://www.cumtp.com> E-mail: cumtpvip@cumtp.com
印 刷 徐州中矿大印发科技有限公司
开 本 787×1092 1/16 印张 10 字数 190 千字
版次印次 2016 年 11 月第 1 版 2016 年 11 月第 1 次印刷
定 价 30.00 元

(图书出现印装质量问题,本社负责调换)



前 言

随着全球定位系统(Global Position System, GPS)设备、射频识别(Radio Frequency Identification, RFID)、无线传感器网络(Wireless Sensor Network, WSN)、卫星和无线通信等技术的快速发展,推动了基于位置的服务(Location Based Service, LBS)快速发展。而LBS的基础就是移动对象轨迹数据挖掘。近年来,LBS的需求不断增长,也为移动对象数据挖掘带来了许多新的挑战 and 机遇,如移动对象数据挖掘与机器学习的结合、移动大数据、基于位置的社交网络等。本书以移动对象轨迹数据为研究对象,以移动对象的活动模式发现为研究的主要目标。

本书从四个层面的理论研究和实践来分析移动对象的数据挖掘方法,期望能够灵活、全面地发现移动对象的周期活动。四个层面包括:在微观层面上,从移动对象的轨迹数据中分析对象的活动特征;在宏观分析层面上,以移动对象轨迹数据为基础,对移动对象进行活动发现,借助于现有的方法对移动对象的兴趣活动进行发现,并推荐对象潜在的兴趣活动,结合移动对象的兴趣相似近邻,发现对象共同感兴趣的活動路径;在模式发现层面上,以对象的活动为基础,通过对移动对象的活动序列进行时间和空间上的多粒度描述,给出对象活动的多粒度周期模式发现方法;在应用层面上,结合本书的研究成果给出相关研究成果的原型系统和领域应用。书中的部分内容包含作者近年来在移动对象数据挖掘方面的一部分成果,如基于结构特征的轨迹聚类、移动对象的活动发现、移动对象的多粒度周期模式发现等。

本书以移动对象的活动轨迹数据为研究对象,以发现移动对象活动中有意义的活动模式为研究目标。在本书知识点规划的基础上,内容共分八章,各章内容组织如下:第1章阐述了移动对象数据挖掘技术的背景及意义,介绍了移动对象数据挖掘研究中面临的挑战和存在的问题,提出了构建移动对象轨迹数据挖掘方法的研究思路和主要研究内容。第2章重点介绍



了移动对象数据挖掘中与本书相关的理论和技术,结合相关技术和国内外研究现状对本书研究内容涉及的技术进行了分析,并对本书的研究内容进行了探讨。第3章从分析典型的移动对象数据挖掘系统入手,论述移动对象数据挖掘系统的需求特点及研究目标;在现有移动对象数据挖掘系统的基础上,构建移动轨迹数据挖掘系统模型,并给出系统的框架结构和系统的分层结构。第4章在充分分析现有移动对象数据挖掘不足的基础上,提出一种基于结构特征的移动对象轨迹数据分析方法;通过构建轨迹结构特征模型,并给出轨迹结构相似度的匹配方法,在轨迹结构相似度的基础上给出轨迹聚类算法来加强轨迹分析的效果。第5章在移动对象轨迹数据的微观分析基础上,提出一种移动对象宏观行为模式分析——移动对象兴趣活动发现方法。首先给出移动对象活动的定义,通过基于密度聚类的方法发现对象已知的兴趣活动,并计算移动对象对相关活动的兴趣度,然后借助于协同过滤技术发现移动对象未知的兴趣活动,在此基础上发现移动对象的兴趣活动路径。第6章在移动对象兴趣活动发现的基础上,结合移动对象的活动在时间、空间上的特点,提出移动对象多粒度周期活动发现方法;在第5章内容的基础上,引入时间粒度的概念,通过对移动对象活动的时间信息进行多粒度描述,对移动对象单活动的多粒度周期以及关联活动的多粒度周期情况进行计算和发掘,并以此来分析移动对象活动的周期模式。第7章以本书介绍的移动对象数据挖掘方法为基础,集成移动对象数据挖掘实验平台,设计并实现了移动对象数据挖掘的原型系统,并以煤矿井下人员定位数据为例,结合煤矿等相关领域的实际情况及其数据特点,给出移动对象数据挖掘在煤矿井下作业人员行为分析上的启发式应用。第8章为总结和展望,在本书介绍的移动对象数据挖掘技术的基础上,从全局角度对移动对象数据挖掘工作进行总结,并指出前期研究过程中产生的有趣问题以及未来可能出现的新研究方向。

中国矿业大学计算机学院夏士雄教授、孟凡荣教授、张磊副教授、牛强教授、周勇教授等对本书的出版给予了极大的鼓励和支持,在此表示深深的感谢。中国矿业大学计算机学院李雯博士生也为本书的付梓提供了全力的协作和密切配合,在此一并表示感谢!

本书的出版得到了江苏省自然科学基金项目(No.BK20130208)、中国矿



业大学优秀青年骨干教师项目以及广西可信软件重点实验室开放基金(KX201613)的资助,在此表示感谢。

数据挖掘的相关技术发展非常迅速,目前已经成为一个广袤的学科。由于笔者才疏学浅,仅略知皮毛,加之精力和时间有限,书中疏漏与错谬之处在所难免,若蒙读者批评指正,不胜感激。

著 者

2016年8月



目 录

1 绪论	1
1.1 背景及意义	1
1.2 技术路线及研究内容	5
1.3 小结	10
2 移动对象轨迹数据挖掘相关技术	11
2.1 移动对象数据挖掘的相关技术	11
2.2 国内外研究现状	14
2.3 本书相关的移动对象数据挖掘工作	18
2.4 小结	23
3 移动对象轨迹数据挖掘系统框架	24
3.1 引言	24
3.2 移动对象轨迹数据挖掘系统框架的提出	24
3.3 移动对象轨迹数据挖掘系统框架	25
3.4 移动对象轨迹数据挖掘的关键技术	34
3.5 小结	41
4 基于结构特征的移动对象轨迹数据聚类分析	42
4.1 引言	42
4.2 轨迹的划分方法	44
4.3 轨迹结构特征与结构相似度计算	46
4.4 索引树的构建	49
4.5 基于结构特征的轨迹片段聚类算法	54
4.6 实验及分析	56
4.7 小结	63



5	基于协同过滤的移动对象兴趣活动发现	64
5.1	引言	64
5.2	问题的提出	65
5.3	移动对象活动热点区域发现	68
5.4	移动对象兴趣活动及路径发现	75
5.5	实验及分析	86
5.6	小结	92
6	基于多粒度的移动对象周期活动发现	93
6.1	引言	93
6.2	周期活动发现的相关定义	96
6.3	多粒度对象活动发现	97
6.4	移动对象周期活动的发现	105
6.5	实验及分析	110
6.6	小结	116
7	原型系统的设计与领域应用	117
7.1	引言	117
7.2	系统结构设计	117
7.3	系统主要模块的构建	120
7.4	原型系统的开发与实现	125
7.5	原型系统在煤矿领域中的启发式应用	129
7.6	原型系统的应用成果展示	132
7.7	小结	134
8	结论与未来展望	135
8.1	总结	135
8.2	未来展望	136
	参考文献	139



1 绪 论

本章首先介绍移动对象数据挖掘的背景及意义,阐述移动对象数据挖掘的技术路线,给出移动对象数据挖掘的主要研究工作和研究思路。

1.1 背景及意义

1.1.1 移动对象数据挖掘的背景

数据挖掘技术是一个从大量数据中提取或挖掘出人们感兴趣的知识的复杂过程,这些知识以隐含的、有效的、新颖的、潜在有用的以及最终可以被理解的模式蕴含在数据之中^[1,2]。1989年8月,在美国底特律召开的第十一届国际人工智能联合会议(IJCAI)的专题讨论会上,数据库中知识发现(Knowledge Discovery in Database, KDD)这个概念被首次提出^[3]。自此,引发了数据挖掘研究的新篇章。经历20多年的发展,数据挖掘的研究已经取得了重大进展,并应用到众多领域。随着研究的深入和新应用需求的层出不穷,越来越多的问题呈现出来,对广大科研工作者提出了更高的要求。随着社会的发展,数据形式变得多种多样,传统基于数值型的数据挖掘和知识发现相关技术已不能满足人们的需求。1994年,在加拿大渥太华举行的GIS国际学术会议上,我国著名学者李德仁教授提出了从GIS数据库中发现知识的概念,并系统分析了空间知识发现的特点和方法,认为它能够把GIS有限的数据库变成无限的知识,并进一步用于精练和更新GIS数据,使GIS成为智能化的信息系统^[4]。经过近20年的发展,国内外对空间数据挖掘已取得了一些成果,但空间数据挖掘的研究无论是在理论研究还是相关软件原型研制等方面都处于发展的初期。

近年来,卫星通信、GPS设备、RFID、无线传感器、物联网通信、视频跟踪等技术的不断发展与广泛应用,使得全球范围内各种大小的移动对象都得到较为准确的定位和有效跟踪。通过这些技术,信号接收设备可以从定位终端上采集到大量移动对象的轨迹数据,这些数据蕴含着非常丰富的信息,而且随着时间



的推移,数据量会变得越来越庞大、复杂,迫切需要研究人员进行进一步的扩展和灵活的分析。由此,使得移动对象数据挖掘这一研究方向应运而生。在2007年的亚太地区知识发现与数据挖掘会议上,世界著名计算机科学家、美国学者J. W. Han教授做了题为《高级数据挖掘技术和应用中的研究前沿》^[5]的特邀报告,报告中将移动对象的数据挖掘作为数据挖掘研究前沿之一。

到目前为止,虽然已经有大量研究是针对定位的移动对象数据展开分析和研究,如对象运动的轨迹数据^[6-8]、轨道交通数据^[9-12]、气候监测数据^[13-18]、动物迁徙数据^[19-24]等一系列定位跟踪数据,但是,针对这些移动对象轨迹数据集的知识发现,很少有系统性的指导和研究方案。此外,由于各种定位和跟踪技术的发展和提高,各行业领域对基于位置的相关服务(Location-based Service, LBS)也显示出巨大的需求^[25-27],这些不断增长的需求对时空数据挖掘技术,尤其是移动对象数据挖掘技术提出了更高的要求。而时空数据挖掘(移动对象数据挖掘)的发展尚跟不上应用需要的步伐,移动对象数据挖掘无论在理论研究还是相关数据分析软件原型研制等方面,都还处于发展的初级阶段,需要进一步深入、系统地研究。因而,移动对象数据挖掘已经成为数据挖掘学科中的主要挑战之一^[28, 29]。

1.1.2 移动对象数据挖掘概述及研究意义

移动对象的数据挖掘主要是通过数学、统计学、人工智能、机器学习等一系列数据挖掘的方法,从移动对象的历史活动数据——轨迹数据中展开分析和挖掘,期望发现新颖的、有价值的信息^[30]。轨迹数据作为移动对象的历史活动数据,在某种程度上能够体现移动对象的属性、状态、行为等内外部特征,此外,还能够体现外部环境变化对移动对象活动的影响等。随着世界的移动程度不断增强、跟踪技术的不断发展,很多政府部门、企业、科研机构等对移动对象的行为分析予以密切的关注,同时,移动对象行为的监测和分析在越来越多的领域内得到应用和发展,如感知矿山与智能矿山管理、智能交通管理、气候监测与天气预报、动物迁徙与环境监测、国家安全与军事安全等。为了更好地分析移动对象的活动特征,把握移动对象的活动趋势及其环境特征,就必须对移动对象的轨迹数据进行系统、有效的分析和挖掘。

移动对象数据挖掘作为一个新兴的数据挖掘学科分支,在工业生产、轨道交通、环境监测、国防安全等诸多领域具有巨大的现实意义和应用价值。通过对移动对象数据挖掘的理论和进行深化和发展,并应用这些技术服务于社会的生产发展,已经成为推动现实世界不断发展的主要驱动力。

(1) 工业生产、企业应用领域



以典型的煤矿企业生产为例,由于安全生产的需要,煤矿井下人员定位系统在煤矿的普及率越来越高^[31]。在煤矿人员定位系统长期运行后,会产生大量的煤矿井下人员活动的轨迹数据,对这些海量的历史轨迹数据进行分析和挖掘,能够获得煤矿井下作业人员的活动规律以及工作状态,这些数据可以用来指导有关煤矿安全生产等方面的方针、政策的制定,具有重要的实际意义^[32,33]。此外,由于建设感知矿山、智能矿山的需要,煤矿生产管理部门会在矿井及生产区域分布相当数量的传感器,这些传感器在其生命周期内不停地收集人员或设备的状态数据。从这些感知数据中可以分离出体现移动对象行为的轨迹数据,结合数字矿山^[34]的 GIS 数据,可以在现代数字化矿山环境中进行移动对象行为模式分析以及煤矿周边地区(如塌陷区、矸石山等)的生态环境变化(通过对移动生物的跟踪和行为分析)监测。

(2) 轨道交通应用领域

很多城市的公共交通工具以及其他一些运输车辆都配备了 GPS 和相关的定位跟踪设备,这些设备以一定的频率向这些车辆的管控中心定时发送车辆的位置信息。根据这些按时间顺序连接起来的定位信息点,车辆管理部门就可以得到车辆的运行轨迹、运动模式^[35]。应用这些 GPS 定位数据以及对驾驶人员行为和车辆运行状况的分析,不仅可以保证对车辆的安全和行驶进行有效调度,还可以借助相关车辆运行信息系统^[36],对道路状况^[37]、驾驶人员的行为特点以及交通流量等信息进行深入的分析,并将分析结果作为车辆管理单位对驾驶人员的驾驶行为进行评估的依据,以及为城市的智能交通系统提供决策支持^[38]。

(3) 环境与气象监测领域

最常见的是通过对气流的定位数据展开分析,进行气候监测与天气预报,如文献^{[39][40]}。通常气象监测机构通过跟踪大气流的运行来记录气流的位置、速度、气压、风力等信息。通过对这些数据的分析,能够发现气流活动的规律及发生的异常情况,从而实现对气候监测和异常天气情况^[13,16]的预报。

(4) 生物活动研究领域

通过对活动动物进行跟踪定位,可以产生大量的轨迹数据,对这些数据展开分析,可以发现活动生物个体的习性、迁徙特点等,如动物的运动除了其自身生活的需要(如迁徙、觅食、休眠等)外^[41],还会受到外界环境的影响。众所周知,在地震前夕一些动物的异常行为频繁发生;受到污染的河水会影响水生生物的生存,导致它们向没有污染或污染较少的区域移动等^[42,43];此外,还可以发现活动生物群体活动的特点。不仅野外的活动生物可以作为研究的对象,



在城市生活的人群也可以作为研究对象,比如:Y.Zheng等^[44]通过对城市人群的活动轨迹进行研究,能够发现他们特殊的地理相关喜好,为进一步研究他们的活动提供了依据。

(5) 隐私保护与安全应用领域

由于大量的移动对象可以被跟踪,如果这些移动对象的轨迹数据被泄露的话,那么将给移动对象活动的安全带来极大的危害^[45, 46]。由此,对移动对象的轨迹数据进行必要的隐私保护,可以尽量减少移动对象相关活动的信息泄露,从而达到减少或减轻由此给移动对象带来的危害^[47]。

此外,还有诸多领域在定位服务、移动对象行为分析等相关服务方面对移动对象轨迹数据挖掘提出了更多的需求^[48]。

近10多年来,数据挖掘技术,特别是以知识发现为基础的数据挖掘理论、方法发展迅速,使用数据挖掘技术从海量数据中抽取对用户有用的知识逐渐得到领域专家、科研工作者、工程技术人员等的充分重视。移动对象行为模式的分析和研究将致力于采用数据挖掘的方法,以移动对象历史轨迹数据为研究对象,以发现移动对象的特征、活动方式等诸多有价值的信息为目标,为现代工业生产、科学研究提供帮助。

1.1.3 移动对象数据挖掘中面临的挑战和存在的问题

在分析移动对象轨迹数据的特征及现有移动对象轨迹挖掘理论、方法、技术的基础上,总结移动对象数据挖掘相关研究在移动对象的轨迹数据挖掘方面存在的问题和面临的挑战,主要包括如下几个方面^[49]。

(1) 高效分析。轨迹数据是对象移动的历史记录,也是包含对象运动特征的元数据。轨迹是由若干采样点组成的位置序列集合,现有的研究往往过分关注于轨迹的整体性,而忽略轨迹的内部结构和轨迹中具有重要价值的局部特征,造成轨迹分析结果的不准确、分析效率不高。

(2) 全面发现。在移动对象轨迹挖掘过程中,移动对象兴趣区域发现具有非常重要的意义。在移动对象轨迹数据库中,一个移动对象只有一条轨迹,随着时间的推移,轨迹会变得越来越长、越来越复杂。传统的研究采用基于密度的方法对轨迹进行分析,确定移动对象的停留区域,然而通过这种方式发现的移动对象会有较多的冗余且不够全面。因为这些停留区域有可能是一个大的区域的一部分;另外,针对单一对象发现的兴趣区域并不全面,无法发现其潜在的兴趣区域。因此,这种方式无法发现更为全面的移动对象兴趣区域。

(3) 灵活抗噪。移动对象的采样轨迹是通过 $(\langle x, y \rangle, t)$ 的形式表示的。



由于多种原因的存在,会造成轨迹数据在定位上的偏差,如果以移动对象原始轨迹进行对象活动的相关模式发现,容易带来较低的精度。此外,由于移动对象的活动具有一定的随意性,在对象的活动序列中会出现偏差,传统的对象活动模式发现方法会发现较多的冗余信息。而在本书中,通过基于活动的对象周期模式发现,不仅可以避免轨迹定位问题带来的不足,还可以通过设置噪声扰动系数,来避免噪声对周期发现结果的干扰。

(4) 领域适应。现有的研究没有以某特定领域展开分析,缺乏说服力。由于特定领域内的移动对象轨迹更具有针对性、特殊性,因此,结合领域知识,对移动对象轨迹进行分析和挖掘更具真实性。在特定领域内,移动对象有不同的角色,不同的角色具有不同的活动环境和移动路径,而这些活动环境和移动路径在不同的背景下具有不同的意义,需要进行语义标注。因而,在对移动对象轨迹进行分析时,需要考虑移动对象的角色和移动轨迹的语义信息,以此来增强领域的适应性。

基于以上分析,本书针对移动对象的历史移动数据——轨迹数据,重点研究移动对象轨迹数据挖掘中的四个重要问题:① 研究能够有效地从轨迹结构角度分析移动对象周期行为模式挖掘的框架,建立移动对象轨迹数据库,解决轨迹数据的存储和重构问题,根据不同的任务对轨迹进行建模和分析,从复杂轨迹的划分、结构抽取、表示等方面构建基于结构特征的移动对象轨迹数据分析方法,并在此基础上开发移动对象轨迹数据挖掘系统框架,实现移动对象轨迹数据挖掘软件的基本原型。② 研究如何全面发现移动对象的兴趣区域的模型和方法,能够更好地从复杂轨迹中发现移动对象停留区域,并能够识别停留区域与兴趣区域的关系,以及移动对象已知的兴趣区域发现以及潜在的兴趣区域发现。③ 研究从多粒度发现移动对象周期行为模式的方法,能够以移动对象的兴趣区域为输入,发现在不同时间粒度下移动对象的周期模式,以及兴趣区域的关联模式。④ 构建移动对象轨迹数据挖掘原型系统,结合相关领域的应用类型和数据特点,研究如何利用本书的成果以及在相关领域内对移动对象活动进行发现和分析的问题,并给出应用成果展示。

1.2 技术路线及研究内容

本书以移动对象轨迹数据为研究对象,以移动对象活动模式发现为最终目标,研究灵活、全面的移动对象轨迹数据挖掘的相关理论和方法。本书从四个层面理论研究和实践分析移动对象的数据挖掘方法,期望能够灵活、全面地



发现移动对象的周期活动。四个层面包括：在微观层面上，从移动对象的轨迹数据中分析对象的活动特征；在宏观分析层面上，以移动对象的轨迹数据为基础，对移动对象进行活动发现，并借助于现有的方法对移动对象的兴趣活动进行发现，推荐移动对象潜在的兴趣活动，结合移动对象的兴趣相似近邻，发现对象共同的感兴趣的活动的路径；在模式发现的层面上，以移动对象的活动为基础，通过对移动对象的活动序列进行时间和空间上的多粒度描述，给出对象活动的多粒度周期模式发现方法；在应用层面上，结合本书的研究成果，给出相关研究成果的原型系统和领域应用。综上所述，本书主要研究内容包括：① 基于结构特征的轨迹数据分析；② 基于协同过滤的移动对象兴趣活动发现；③ 基于多粒度的移动对象周期模式发现；④ 移动对象轨迹数据分析原型系统的设计、实现与领域应用。

1.2.1 基于结构特征的轨迹数据分析

移动对象的轨迹数据，从直观上看，是由若干采样点按照时间顺序组成的位置序列，从逻辑结构上看，是由对象标识、采样位置、采样时间等信息组成的多维数据。在传统关于轨迹数据分析的研究中，存在两方面的不足：一是不够全面，二是过分全面。这个看似矛盾的说法实质上是指：① 在分析轨迹数据时往往仅仅从采样点本身出发，而忽略了轨迹的整体特征，打破了轨迹的内在连续，无法从全局角度分析轨迹的特征和运动趋势，所以这个不足之处称为不够全面。② 由于移动对象的轨迹具有局部的相似性和全局的不相似性，传统的研究往往以整条轨迹为单位进行分析研究，没有看到轨迹的显著特征往往集中于轨迹的某一部分，另外，对于很长而且很复杂的轨迹，不利于发现其蕴含的信息，这种情况称之为过分全面。任何一种情况的发生都容易造成轨迹分析的不准确。③ 现有的空间索引结构大多以轨迹的位置点进行存储并用最小外接矩形(Minimum Bounding Rectangle, MBR)对移动对象的活动轨迹进行存储，容易带来空间上的重叠性，而且不利于表示轨迹的特征，如轨迹的速度、方向、转角等相关的特征信息。

根据现有研究在轨迹分析方面的不足之处，本书提出基于结构特征的移动对象轨迹相似度计算方法，该方法首先计算轨迹的每个转角，在转角较大的采样点处(也称为突变点)对轨迹进行划分，使每一段的轨迹结构较为平稳。接下来，引入轨迹结构的概念对轨迹进行建模，并抽取方向、速度、转角、距离等内外部特征对轨迹进行表示。然后，给出结构相似度计算函数来分析轨迹内外部特征，通过设置不同的阈值，可以根据应用领域的需要任意调节各种特征的敏感程度，使结果更加贴合实际情况。本书根据转角将轨迹划分成若干



轨迹段,并通过计算轨迹段的结构相似度来判断轨迹的匹配程度,进而完成轨迹分析。真实数据的实验结果表明:这种算法效率高,还可以通过不同参数灵活调整特征的敏感程度,使轨迹的分析结果更具有实际意义。

简言之,本书就是通过对移动对象轨迹数据进行建模并抽取其结构特征,分析其内外部的特征变化及运动趋势,通过提出轨迹结构相似度的匹配方法,进行轨迹特征比较。在进行轨迹比较前,先根据内部的变化情况,将轨迹划分成若干轨迹片段,使每个轨迹片段都具有较为平稳的结构特征。将基于结构相似度的轨迹数据分析方法应用于轨迹的聚类与异常检测。

1.2.2 基于协同过滤的移动对象兴趣活动发现

通过对轨迹数据中的 Stop-Move(STPM)点进行识别,可以发现轨迹数据的运动特征,这些 STPM 点反映了移动对象在过去一段时间内的运动形式以及其停留区域。通过对 Stop 点进行聚类可以发现移动对象停留时间相对较长、面积相对较大的区域,对于这些区域,我们称之为移动对象的兴趣区域。由于多个移动对象之间往往具有相似的经历、相似的特征、相近的兴趣爱好,因此,他们的兴趣区域往往表现为地理空间上的一些重叠区域。这些重叠区域就是他们共同关注的区域,反映了他们在空间爱好上的某些特征的相似性,而这些重叠区域的出现次数、发生时刻、停留时间等是衡量移动对象之间相似程度的重要指标。通过设计合适的算法,可以计算他们之间的相似程度,在相似程度较高的对象之间形成一种兴趣区域推荐机制,采用最近优先的推荐方法进行兴趣推荐。通过这种方式,既可以发现移动对象已知的兴趣区域,还可以知道他们的潜在兴趣区域。例如,设 A 和 B 是相似程度较高的两个对象,如果 A 访问的区域 B 没有访问,并且这些区域对 B 又有重要的意义,则可以将这些区域推荐给 B,形成 B 的潜在兴趣区域。此外,在对象的兴趣活动区域的基础上,形成移动对象的活动序列,并且根据其近邻的活动路径,发现对象感兴趣的活動路径。

1.2.3 基于多粒度的移动对象周期模式发现

周期性是移动对象轨迹数据的重要特性之一,周期模式挖掘是移动对象模式分析的基本应用之一。本书首先对移动对象轨迹数据进行基于密度的聚类,发现其经常停留的热点区域(Hot region);接下来,引入时空多粒度对象活动发现算法,对移动对象的活动进行多粒度表示;然后,对移动对象的多粒度活动进行遍历,并标记移动对象对于热点区域的每一次访问为移动对象的一个活动,每一个活动都记录相关活动区域、发生时刻、停留时间等信息,形成



对象的活动序列。然后,利用多粒度时间的表示方式,给出不同时间粒度之间的转换关系、包含关系等进行描述和定义,结合时间粒度的概念对活动发生时刻进行多粒度的描述,在多种时间粒度下对移动对象的周期模进行描述。在此基础上,对移动对象的单活动周期关联关系、移动对象的多活动周期关联关系进行进一步分析,通过借助如最大子模式树(Max-subpattern tree)等一些存储和数据表示的结构,研究移动对象的周期模式挖掘方法,从非常全面的角度分析移动对象的周期模式。

1.2.4 移动对象轨迹数据分析原型系统

为了能够深入地对移动对象轨迹数据展开分析,同时也为了丰富和发展移动对象数据挖掘技术,需要根据移动对象轨迹数据挖掘的相关需求以及移动对象数据挖掘的应用需求,构建移动对象轨迹数据挖掘的原型系统。

针对本书给出的移动对象轨迹数据挖掘的研究内容,需要从以下几个角度对原型系统进行设计:① 轨迹数据的收集、清洗与预处理。由于现有的定位采集设备多种多样,如:GPS、无线传感网络、RFID等,而且不同的定位设备采集的轨迹数据具有异构性,因此需要一个专门的基础模块用来对轨迹数据进行标准化处理,便于进一步的分析。② 针对本书的第一个研究内容,设计移动对象轨迹数据结构分析模块,以该研究的主线为设计思路,用于从微观层次发现移动对象的运动特征。③ 针对本书的第二个研究内容,设计移动对象的兴趣活动发现模块,用于从宏观层次发现对象的兴趣活动。④ 针对本书的第三个研究内容,设计移动对象的周期活动发现模块,用于从模式发现的角度对移动对象的活动模式进行分析。

煤矿领域为了生产发展的需要,往往在煤矿井下及矿区生产环境中布置了大量的传感器,用来获取移动对象的定位数据,辅助煤矿安全生产管理。在这些应用中,最为常见且最为重要的是煤矿井下人员定位系统,通过在煤矿井下巷道、采掘工作面等作业环境下布置传感定位设备,记录煤矿作业人员的活动轨迹。这些定位的活动轨迹数据包含了大量的信息:如作业人员的活动状态、作业人员的出勤信息等,迫切需要深入分析。因此,借助移动对象轨迹数据挖掘原型系统发现煤矿井下作业人员的活动规律和特点,是一个很重要的实践和创新。针对这一点,本书拟给出原型系统在煤矿领域的启发式应用,期望将移动对象数据挖掘方法与领域数据结合,为工业生产等实际应用领域服务。

此外,为了便于原型系统的进一步扩充和深化,在原型系统的设计过程中,我们拟采用面向对象的模块化设计思想,降低各模块之间的耦合度,同时增强代码的可复用性。



基于上述的研究技术路线和主要研究内容,本节给出本书研究的技术路线图,如图 1-1 所示。在图中,我们可以看出本书进行的研究内容及研究成果。本书的研究采用层次递进的方式逐步深入展开,其中移动对象的轨迹元数据管理和轨迹数据的预处理作为研究的基础工作,不在本书中展开讨论。本书从基于结构特征的轨迹分析方法开始,对移动对象的活动轨迹数据展开深入的分析和研究,通过研究出的有效方法为移动对象活动发现提供方法理论。基于协同过滤的对象活动发现方法,主要在轨迹数据上进行对象的活动发现,并借助协同过滤方法,发现对象的兴趣活动,结合与移动对象相似的近邻,进一步发现对象的兴趣活动路径。基于多粒度的移动对象周期模式发现,则在对象兴趣活动发现的基础上,对移动对象的活动进行多粒度描述,并形成对象活动的多粒度序列,在此序列的基础上发现对象活动的多粒度周期模式。最后,将本书的研究成果通过移动对象周期行为挖掘原型系统体现出来,并结合煤矿领域的特点,将本书的研究成果应用于人员定位轨迹数据,给出煤矿领域的启发式应用。

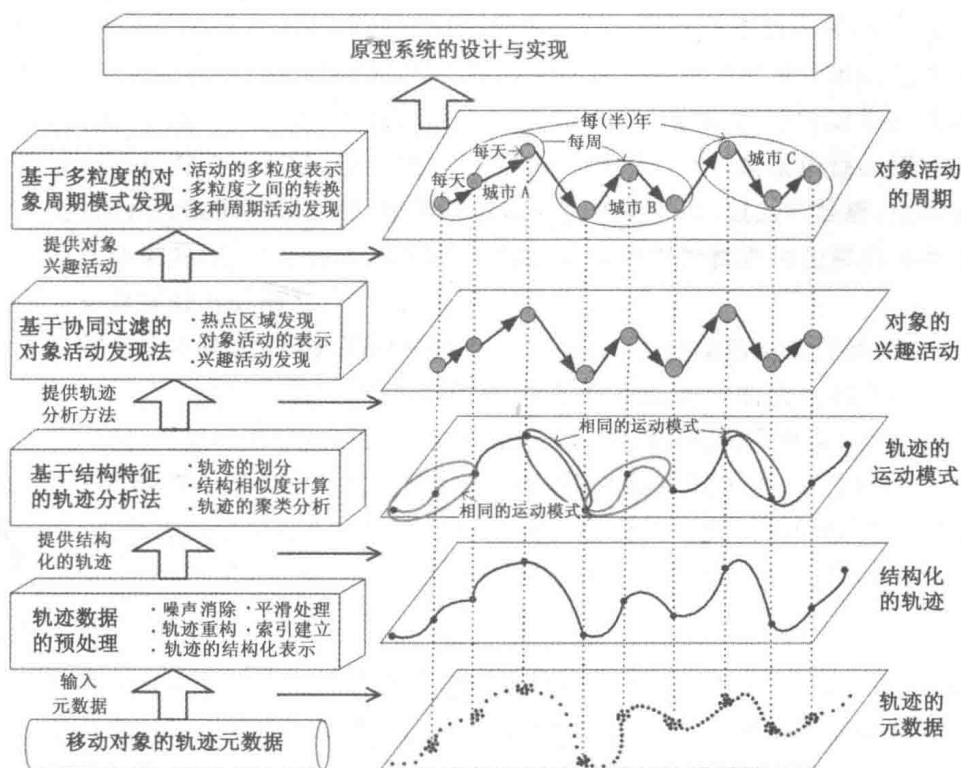


图 1-1 研究技术路线图