

基于交通CPS的流式数据聚类及 演化趋势发现方法研究

S

STUDY ON CLUSTERING AND EVOLUTIONARY TRENDS DISCOVERY OF MULTIPLE DATA STREAMS
IN TRANSPORTATION CYBER PHYSICAL SYSTEMS

桑春艳◎著



重庆大学出版社

基于交通CPS的流式数据聚类及

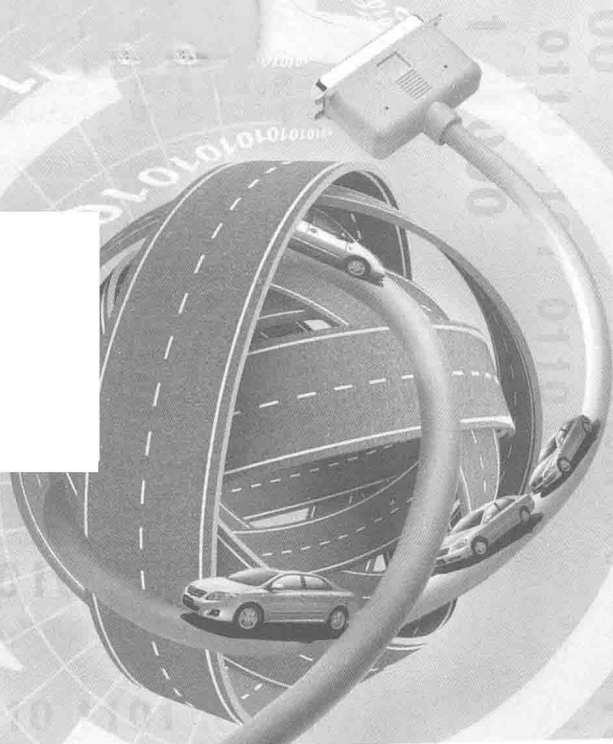
演化趋势发现方法研究

J

IYU JIAOTONG CPS DE LIUSHI SHUJU JULEI JI
YANHUA QUSHI FAXIAN FANGFA YANJIU

桑春艳◎著

重庆大学出版社



内容提要

信息物理系统(Cyber Physical Systems, CPS)的提出及应用为解决交通系统中存在的问题提供了新思路。将 CPS 技术应用于交通系统,一方面,可广域多维地获悉表征交通物理系统实时状态的信息,为获悉交通物理系统实时状态和运行规律提供了重要的信息来源;另一方面,通过对所获取的海量交通数据的及时分析和有效处理,进而为交通物理系统的全面协调和实时优化提供新的依据。

本书在现存流式数据聚类方法研究的基础上,研究表征交通物理系统状态广域多维的交通多流式数据的聚类分析及交通多流式数据的演化趋势发现方法。针对交通流式数据的周期演化特性,受启发于联合聚类以及基于矩阵分解聚类的思想,提出了基于低秩近似矩阵分解的多流式数据进化聚类算法 EC-NMF。针对交通系统中流式数据随空间演化的纵向传播特性,提出了基于非负矩阵三分解的交通多流式数据联合聚类框架 STClu。为揭示交通流式数据之间随时空的演化特性,提出了基于聚类思想的交通多流式数据演化趋势发现方法。

本书适合有一定数据分析基础的学生、研究者阅读。

图书在版编目(CIP)数据

基于交通 CPS 的流式数据聚类及演化趋势发现方法研究/桑春艳著. —重庆:重庆大学出版社,2017.5
ISBN 978-7-5689-0317-2

I. ①基… II. ①桑… III. ①数据处理—聚类分析—研究 IV. ①TP274

中国版本图书馆 CIP 数据核字(2016)第 308183 号

基于交通 CPS 的流式数据聚类及 演化趋势发现方法研究

桑春艳 著

策划编辑:何 梅

责任编辑:李定群 版式设计:何 梅

责任校对:贾 梅 责任印制:赵 晟

*

重庆大学出版社出版发行

出版人:易树平

社址:重庆市沙坪坝区大学城西路 21 号

邮编:401331

电话:(023) 88617190 88617185(中小学)

传真:(023) 88617186 88617166

网址: <http://www.cqup.com.cn>

邮箱: fxk@cqup.com.cn (营销中心)

全国新华书店经销

重庆紫石东南印务有限公司印刷

*

开本:720mm×960mm 1/16 印张:9.75 字数:135 千

2017 年 5 月第 1 版 2017 年 5 月第 1 次印刷

ISBN 978-7-5689-0317-2 定价:39.00 元

本书如有印刷、装订等质量问题,本社负责调换
版权所有,请勿擅自翻印和用本书
制作各类出版物及配套用书,违者必究

前言

交通需求的持续增长使得交通拥堵、环境污染、交通安全等问题日趋严重。如何运用科学的方法准确认知交通物理系统的演化规律,对现有路网资源进行优化配置,提高现有道路资源的利用效率,为缓解交通拥堵等问题已成为目前亟待解决的问题。信息物理系统(Cyber Physical Systems, CPS)的提出及应用为解决交通系统中存在的问题提供了新思路。将CPS技术应用于交通系统,一方面,可广域多维地获悉表征交通物理系统实时状态的信息,为获悉交通物理系统实时状态和运行规律提供了重要的信息来源;另一方面,通过对所获取的海量交通数据的及时分析和有效处理,进而为交通物理系统的全面协调和实时优化提供新的依据。

将大量的感知设备布设于道路交通系统中用以实时监控道路交通状态,进而获取表征交通物理系统的实时状态信息,通过分析这些

信息并将处理结果反馈于交通物理系统中,进而实现交通物理系统的实时协调和全面优化,体现了 CPS 的典型特征。在交通物理系统中,由 GPS、RFID、感应线圈等不同感知设备所产生的交通数据以流的形式不断涌现。通过对监控道路交通实时状态的交通流式数据的分析,不仅可对道路交通状态进行评价和预测,还可解析交通状态演变的机理,掌握交通物理系统的运行规律。

本书通过对交通流式数据的特点及特性的分析,在现存的流式数据聚类方法研究的基础上,研究表征交通物理系统状态广域多维的交通多流式数据的聚类分析及交通多流式数据的演化趋势发现方法。

首先,从 CPS 的角度分析交通流式数据的特点及特性。为探索基于 CPS 的交通流式数据的分析及处理方法,对表征道路交通状态的常用参数进行了描述,总结了交通流式数据的特点。基于固定检测器所采集的道路实时状态信息,对交通流式数据的周期演化和纵向传播特性进行分析。

其次,为发现交通多流式数据之间的关联关系,结合交通流式数据的周期演化特性,提出交通多流式数据的进化聚类分析方法。为解决交通多流式数据聚类时的高维问题,受启发于

联合聚类以及基于矩阵分解聚类的思想,提出了基于低秩近似矩阵分解的多流式数据进化聚类算法 EC-NMF。EC-NMF 算法充分利用流形与低秩结构来学习非负数据的有效表示,分别在数据空间和特征空间中构建基于近邻的数据图和特征图来反映它们各自的几何流形结构。为保持聚类结果随时间变化的平滑性,EC-NMF 算法考虑了随时间滑动的历史聚类结果的信息。

再次,结合具有上下游关系的交通流式数据之间的纵向空间传播特性,基于联合聚类多个相关类型数据的研究现状,提出了交通多流式数据的联合聚类算法。为了能够更客观地分析多交通流式数据之间的关联关系,结合交通系统中流式数据随空间演化的纵向传播特性分析,提出基于非负矩阵三分解的交通多流式数据联合聚类框架 STClu。

最后,为揭示交通流式数据之间随时空的演化特性,进一步获悉多个相似断面之间的交通状态随时间演化的特性,提出了基于聚类思想的交通多流式数据演化趋势发现方法。该方法以单条流式数据为单位的聚类问题转化为多流式数据的图聚类模型。根据交通流式数据之间的滞后相关性特征,给出了基于滑动窗口的交通流式数据的滞后相关性计算方法。基于谱

图理论的相关思想,提出了基于滞后相关的交通多流式数据的聚类算法 ICMDs。为通过分析不同时刻交通多流式数据的聚类结果,获悉交通流式数据的演化趋势,提出了基于 ICMDs 算法的交通多流式数据演化趋势发现算法 TEEMA。

限于本书著者的学识水平,书中疏漏与不妥之处在所难免,恳请读者批评指正。

著 者

2016 年 8 月

目 录

第 1 章 绪 论	1
1.1 研究背景及意义	2
1.1.1 研究背景	2
1.1.2 研究意义	5
1.2 国内外研究现状	7
1.2.1 CPS 的发展现状	7
1.2.2 T-CPS 的相关研究	11
1.2.3 交通数据的分析方法	13
1.2.4 流式数据的聚类分析	16
1.3 主要研究内容	21
1.4 本书的组织结构	24
第 2 章 基于 CPS 的交通流式数据的特点 及特性分析	26
2.1 交通状态描述	26
2.2 基于 CPS 的交通流式数据的特点	30
2.3 交通流式数据的特性分析	31

2.3.1 交通流式数据的周期演化特性分析	31
2.3.2 交通流式数据的纵向传播特性分析	39
2.3.3 交通多流式数据的相似性演化特性分析	44
2.4 本章小结	46

第3章 基于周期演化特性的交通多流式

数据进化聚类算法	47
3.1 引言	47
3.2 相关工作	49
3.2.1 非负矩阵分解算法	49
3.2.2 图正则约束的非负矩阵分解	51
3.3 基于周期特性的交通多流式数据进化聚类	52
3.3.1 问题描述	52
3.3.2 基于周期特性的交通多流式数据聚类建模	53
3.3.3 迭代更新	54
3.4 算法描述及其分析	57
3.4.1 EC-NMF 算法描述	57
3.4.2 收敛性分析	58
3.4.3 复杂度分析	61

3.5 仿真实验及结果分析	62
3.5.1 比较算法及评估方法	62
3.5.2 合成数据集上的实验结果及分析	63
3.5.3 实测数据集上的实验结果及分析	68
3.5.4 参数选择	71
3.6 本章小结	74

第 4 章 基于纵向空间传播特性的交通多

流式数据联合聚类分析

4.1 引言	75
4.2 相关工作	77
4.2.1 联合聚类	77
4.2.2 基于 NMTF 的联合聚类	78
4.3 基于时空特性的交通多流式数据联合 聚类模型	79
4.3.1 问题描述	80
4.3.2 基于时空特性的交通多流式数据 聚类建模	81
4.3.3 迭代更新	82
4.4 STClu 算法描述及其分析	84
4.4.1 算法描述	84
4.4.2 复杂度分析	86

4.5	仿真实验及结果分析	87
4.5.1	比较算法及参数设置	87
4.5.2	合成数据集上的实验结果及分析	87
4.5.3	实测数据集上的实验结果及分析	93
4.6	本章小结	99
第5章	基于谱图理论的交通多流式数据 演化趋势发现算法	100
5.1	相关工作	100
5.2	问题描述	102
5.3	交通多流式数据的滞后相关性度量	103
5.4	交通多流式数据的演化趋势发现 算法	106
5.4.1	多流式数据的统计分析	106
5.4.2	交通多流式数据的演化趋势发 现算法	108
5.5	仿真实验及结果分析	110
5.5.1	比较算法及度量指标	110
5.5.2	ICMDS 算法的有效性	111
5.5.3	TEEMA 算法的可扩展性	115
5.6	本章小结	117

第 6 章 总结与展望	118
6.1 本书的主要工作	118
6.2 后续工作展望	121
参考文献	122

第 1 章

绪 论

如何运用科学的方法准确认知交通物理系统的运行规律,对现有路网资源进行优化配置,提高现有道路资源的利用效率,以缓解交通拥堵等问题已成为目前亟待解决的关键问题。信息物理系统(Cyber Physical Systems, CPS)的提出为交通系统的全面协调和实时优化提供了新的思路。将大量的感知设备应用于交通物理系统中,使得能够表征交通实时状态的数据以指数级的速度实时增长。例如,由 GPS、RFID、感应线圈等不同感知设备所产生的交通数据以流的形式不断涌现。由于监控道路交通状态的交通数据刻画了交通物理系统的动态变化过程,通过对监控道路交通状态的交通流式数据的分析,不仅可对道路交通状态进行评价和预测,还可解析交通状态演变的机理,掌握交通物理系统的运行规律。

1.1 研究背景及意义

1.1.1 研究背景

(1) 科学地认识交通物理系统的演化规律已成为目前亟待解决的问题。建立绿色、高效、安全的交通系统是交通参与者追求的目标。然而,随着交通需求的持续增长、汽车保有量、交通流量迅速增加,使得交通拥堵、交通安全、环境污染等问题日趋严重。

1) 交通拥堵问题日趋严重

交通拥堵的加剧,使得交通延误增大、行车速度降低、经济损失严重等^[1-4]。道路基础设施的建设在一定程度上缓解了交通拥堵问题,但没有得到实质性的改观。如果缺乏科学理论的指导和先进技术的支撑,仅依靠增加道路基础设施的手段已经不能从根本上解决交通拥堵等问题。

2) 环境污染问题日渐增加

交通拥堵情况下,车辆长时间处于怠速、低速、急加速、急减速状态。车辆排放如 SO_2 , NO_x , CO , CO_2 等有害物质的尾气,对环境的污染极其严重。现有的研究表明,车流处于“时走时停”状态时,直接导致原油消耗,占世界总消耗量的 20%^[3-5]。由此所导致的环境污染和能源消耗等问题日益严重。

3) 交通安全问题形势严峻

随着交通需求的日益增长,汽车保有量的不断增加,交通安全问题形势严峻。据公安部资料统计显示,2007 年,我国共发生公路交通事故 327 209 起,造成 81 649 人死亡,380 442 人受伤;其中高速公路共发生交通事故 12 364 起,造成 6 030 人死亡,14 628 人受伤。2007 年我国高速公路每百千米发生交通事故 23.07 起,同期所有公路每百千米发生交通事

故9.16起;高速公路事故每百千米受伤 27.29 人,同期所有公路的平均事故受伤为 10.65 人/100 km。

为此,以先进技术支撑交通物理系统的发展,用科学的方法准确认知交通物理系统的运行模式,充分有效地利用现有的路网资源,以缓解交通拥堵、环境污染、交通安全等问题已成为目前亟待解决的问题。

(2) CPS 的研究与发展为交通物理系统的协调与优化提供了新的思路

CPS 的提出^[6-10],将飞速发展和日益成熟的计算、通信、感知、控制等技术将物理系统的行为特征和状态实时、协同、安全、可靠地传输到信息系统中。在信息系统中,通过对所获取的物理系统实时状态的信息进行及时分析和有效处理,进而做出对物理系统的控制与决策。通过网络化的控制设备和执行设备,将信息系统及时的控制与决策方案协同实施,对物理系统进行准确实时的协调与优化。

交通物理系统中内部各要素在时空上的全面协调和有序运行,很大程度上依赖于对各交通物理对象实时状态的全面获悉。将大量的感知设备布设于道路交通状态的实时监控中,通过将所获取的表征交通物理系统实时状态的信息的及时分析和有效处理,并将分析及处理结果反馈于交通物理系统中,进而实现交通物理系统的实时协调和全面优化,体现了 CPS 的典型特征^[11-12]。

因此,CPS 的提出为缓解交通拥堵、监控车辆安全、节能减排等问题提供了新的途径。

(3) 交通信息物理系统的理论研究及实际应用受到了广泛重视

交通信息物理系统 (Transportation Cyber Physical Systems, T-CPS) 的提出^[12],一方面发挥了 CPS 技术在感知、传输、计算、控制等方面的优势,实现对交通物理对象,包括道路交通基础设施、载运工具、交通状态等多方面的全面感知,为获悉交通物理系统实时状态和运行规律提供了重要的信息来源^[13-14]。另一方面,在交通信息系统中,通过对所获取的海量交通信息的及时分析和有效处理,可为准确地认知交通物理系统的演化趋势,掌握交通拥堵等问题产生和传播的特性,为交通物理系统的全面协

调和实时优化提供新的依据。

总之,研究 T-CPS 的相关理论方法具有两方面的作用:一方面通过对交通 CPS 的基础理论、框架结构、系统建模等方面的相关研究,不仅为 T-CPS 的发展提供理论支持,还可为 CPS 相关理论的发展奠定基础;另一方面,作为下一代智能交通系统重要的发展方向,T-CPS 的相关研究为解决交通系统中存在的问题提供理论依据和技术支撑。

(4) 交通海量数据的分析及处理方法需要进一步探索

交通数据分析处理方法的目标是在交通信息系统中通过对不同来源的交通数据的实时分析和有效处理,发现交通物理系统的演化规律,并将处理结果反馈于交通物理系统中,进而为交通物理系统的协调与优化提供理论支撑。

随着交通大数据时代的到来,更深入地了解交通物理系统演化特性和运行规律引起了人们的广泛关注。如图 1.1 所示,将大量的感知设备应用于道路交通状态的实时监控中,由此所产生的表征交通状态的信息刻画了交通物理系统的动态变化过程。

针对海量交通流式数据的研究不仅是交通领域的研究热点问题,也已成为机器学习、数据挖掘、模式识别及统计分析等领域研究的热点^[15-18]。现有的交通数据的分析方法大致可从两个方面进行总结:一方面,为了获悉道路交通的状态,借鉴已有的方法应用于问题的解决中,如交通状态判别、预测等;另一方面,通过交通数据特点的分析,研究适合于此类数据的分析方法,如交通数据的聚类分析方法等。

为充分发挥表征交通物理系统实时状态的海量交通数据的作用,Tang 等^[17]提出了基于聚类思想的交通系统中非典型事件的发现方法。从流式数据的角度,Wei 等^[15]结合交通流式数据的时空特性,提出了两阶段的增量聚类算法。Geisler 等^[13]提出了交通信息系统的评估框架。由于交通系统的复杂多变性,基于 CPS 的交通数据的分析方法的研究涉及的范围和内容较广泛,结合交通流式数据特点的聚类分析还有待进一步探索。

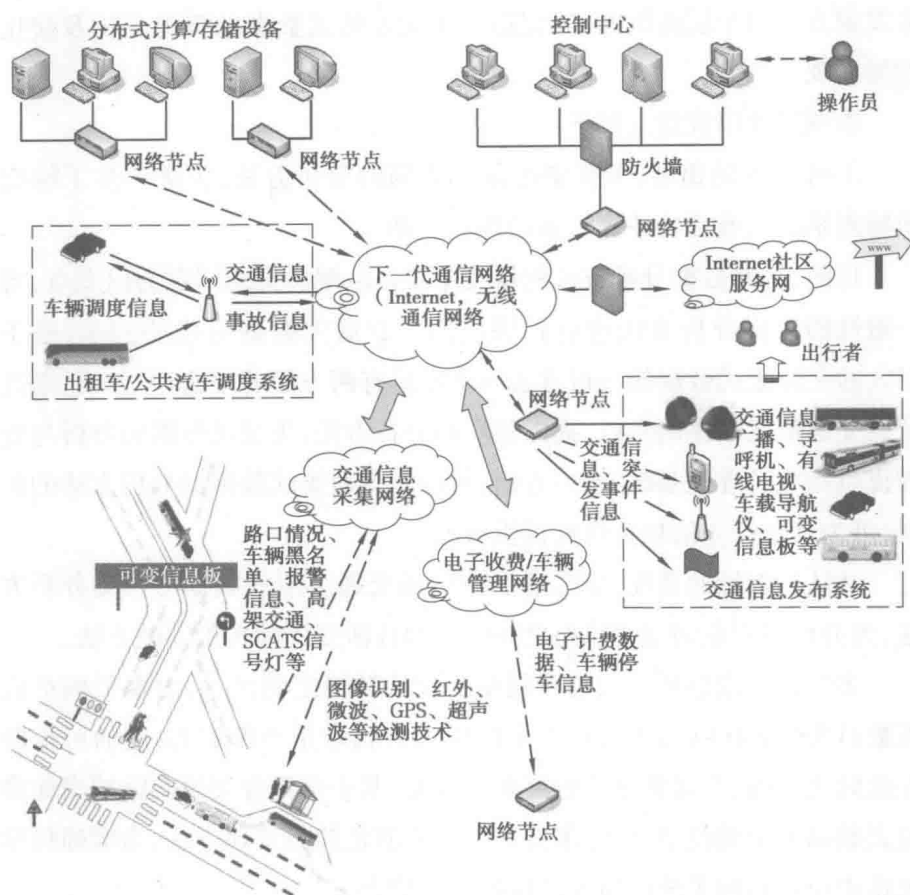


图 1.1 CPS 技术在交通系统中的应用

因此,本书通过对以流的形式实时产生的交通数据特点及特性的分析,以现存的流式数据的聚类分析方法为基础,通过多交通流式数据之间的相关性分析,构建交通多流式数据的聚类模型,研究基于 CPS 的交通多流式数据的聚类分析及演化趋势发现方法,为进一步解析交通状态演变的机理、掌握交通物理系统的演化规律提供新的理论基础。

1.1.2 研究意义

本课题围绕基于 CPS 的交通流式数据的特点,基于流式数据的研究现状及高维数据的聚类分析方法,从多流式数据的相关性分析及演化趋