

交通大数据关键技术研究

顾承华, 张 扬, 翟 希

(上海市城乡建设和交通发展研究院, 上海 200232)

摘 要: 作为数字化时代的新型资源, 大数据成为驱动创新的重要因素, 正在改变着人类社会的发展。大数据对科学研究、经济建设、社会发展和文化生活等各个领域正在产生革命性的影响。以交通数据资源和信息技术为支撑的城市交通决策和服务, 是现代化城市交通的重要标志。近年来, 随着城市交通信息化以及智慧城市建设的不断深入, 各种数据的采集、汇聚已具规模, 交通大数据的产生, 是在大数据技术促进下城市交通信息化发展到一定阶段的必然结果。基于上海市交通信息化建设成就和数据基础, 依托上海市科委《交通大数据应用关键技术与示范》课题研究成果, 对交通大数据框架体系、基础理论、软硬件构建、实践应用等方面进行较为深入的研究和归纳, 并对现代化城市交通管理决策和社会公众服务等方面的需求及发展做出展望。

关键词: 城市交通; 大数据; 实践应用

中图分类号: U491

文献标识码: A

文章编号: 1671-3400(2015)12-0049-04

Research on Key Technologies of Traffic Big Data

GU Cheng-hua, ZHANG Yang, ZHAI Xi

(Shanghai Urban-Rural Construction and Transportation Development Research Institute, Shanghai 200232, China)

Abstract: As a novel resource in digital era, big data has been an important driving aspect of creativity, which is now changing our human society. Big data also brings the revolutionary effects to the areas including scientific research, economic construction, social development, culture and daily life, etc. Urban traffic decision and service supported by traffic data and information technology (IT) are important symbols in modern cities. With the development of traffic informationization and construction of smart city, types of traffic data have been collected and stored in certain scale recent years. Naturally, urban traffic big data rise from the developed traffic informationization in the background of big data technology. On the basis of the development of traffic informationization and stored data in Shanghai, this paper provides the research achievements which involve systematic frames, basic theories, software and hardware, practical applications, etc. It contains some achievements of the key projects from Shanghai Science & Technology Commission. And it also looks forward to the emergent requirements of the application direction of big data in traffic management, decision, and public service as well.

Keyword: Urban traffic; Big data; Practical application

1 大数据与城市交通

大数据概念产生以来, 作为数字化时代的新型战略

资源, 大数据成为驱动创新的重要因素, 正在改变着人类的生产和生活方式。从广义上讲, 大数据有3层内涵: 一是数据量巨大、来源多样与类型多样的数据集; 二是新型的数据处理和分析技术; 三是运用数据分析形成新的价值。目前, 大数据对科学研究、经济建设、社会发展和文化生活等各个领域正在产生革命性的影响^[1]。

随着社会经济发展和城市化进程的加快, 解决好城市交通问题已经成为城市可持续发展的一个重要内容。现代化城市交通管理已经进入以信息化支撑为标志的新

收稿日期: 2015-10-19

基金项目: 本文部分成果来自上海市科委课题研究项目 (12511509600、14QB1403700)

作者简介: 顾承华 (1963-), 男, 上海人, 大专, 高级工程师, 现从事交通信息化研究。

时期,以交通数据资源和信息技术为基础的城市交通决策和服务是现代化城市交通的重要标志^[2-4]。近年来,随着城市交通信息化以及智慧城市建设的不断深入,各种类型数据的采集、汇聚已具有了一定规模,交通大数据的产生,是在先进技术促进下城市交通信息化发展到一定阶段的必然结果。现代化城市交通决策规划、组织管理和公众出行服务等,对大数据技术发展和应用的需求日益迫切。在这样的大背景下,上海从2012年开始进行了较为系统的交通大数据研究,并形成了一系列具有创新性的成果。

2 交通大数据定义与涵盖范围

开展交通大数据研究,需要在交通信息化建设与发展基础上,立足数据资源,认真思考究竟什么是交通大数据,这是首要问题^[5]。

2.1 上海交通大数据基础

就上海来讲,上海市交通综合信息平台(以下简称“交通综合信息平台”或“平台”)建成于2008年3月,由上海市交通信息中心(现上海市城乡建设和交通发展研究院上海交通信息中心)组织建设并负责运行维护,是目前全国规模最大、全面、实时整合、处理全市道路交通、公共交通、对外交通领域车流、客流、交通设施等多源异构基础信息数据资源,实现跨行业交通信息资源整合、共享和交换,为交通管理相关部门进行交通组织管理和社会公众进行交通综合信息服务提供基础信息支持的信息集成系统之一。

平台汇聚了来自上海市交通委路政管理部门、市交警总队、市交通委、机场、铁路、码头等交通管理行业的道路交通、公共交通、对外交通各类交通信息数据共230多项。这些数据来自5.8万组感应线圈、2.5万辆GPS浮动车、300余组车牌识别断面、中国移动手机用户等动态交通数据采集装置,以及对1000多条公交线路、14条轨道交通线路、700多个社会停车场库、2个国际机场、3座铁路客运站的线路分布、实时泊位、航班等动态数据的采集,数据范围覆盖全市域3张路网、公共交通、对外交通等重要交通枢纽和公交线路等。

目前,交通综合信息平台汇聚整合的数据量巨大,主要通过3种形式进行存储和处理。交通综合信息平台移动前端交通信息处理系统每天存储的手机数据、GPS数据量为150GB,以分布式文件存储在各运算节点上,通过分布式流式运算,在内存中进行快速处理,将结果文件存入数据库;交通综合信息平台每天存储的结果数据量为30GB,通过光纤网络将实时、历史交通数据直接存入Oracle数据库的列表,进行索引后实现平台的展示应用和历史分析;交通综合信息平台系统的二级行业

平台,如交警总队、路政管理部门等每天存储的视频信息数据量为450TB,实时存储在二级平台的硬盘里,可供随时调用。

因此,整个交通综合信息平台系统所汇聚整合的数据,形成了大而复杂的数据集,是目前全国规模最大的交通综合信息平台之一。平台也成为上海交通信息交换共享的枢纽,提高了交通信息资源的利用率和业务管理效率,加强了交通管理的协同性,成为现代交通管理的重要手段之一。这些积累形成了开展交通大数据研究的前提和基础。

2.2 交通大数据定义与涵盖范围

交通大数据,本文给出的定义是,由城市交通运行管理直接产生的数据(包括各类道路交通、公共交通、对外交通的线圈、GPS、视频、图片等数据)城市交通相关的行业和领域导入的数据(气象、环境、人口、规划、移动通信手机信令等数据),以及来自公众互动提供的交通状况数据(通过微博、微信、论坛、广播电台等提供的文字、图片、音视频等数据)构成的,用传统技术难以在合理时间内管理、处理和分析的数据集。交通大数据中同时包含了来自交通行业的和交通行业之外的格式化和非格式化数据。

从这个定义可以看到,交通大数据的覆盖范围,主要是交通领域直接产生的数据、公众互动交通状况数据、相关行业数据和重大社会经济活动关联数据这三大类。总体上讲,交通大数据具有了数据量巨大、数据种类多样、蕴含价值丰富的大数据特征。城市交通各类数据的汇集,以及气象、环境等数据的导入,直接导致了交通大数据的数据量巨大;城市交通领域的道路交通、公共交通、对外交通,以及政治、经济、社会、人文等其他领域和行业数据的结构化、非结构化和半结构化,促使交通大数据具有种类多样性;跨地域、跨行业、跨部门的数据,在智慧城市建设背景下,相互交融和关联,可以为管理和服务提供更广阔的视野,从而得到以往无法探知的规律。基于数据关联性研究,对于发掘、利用交通大数据自身的价值,是重点和方向。

2.3 交通大数据体系框架

大数据环境下的交通信息化管理和服务,给数据的处理、分析和应用提出了更高的要求。基于大数据基础理论,在智能交通这个特定领域内,研究相适合的理念、理论、技术、应用等,需要构建一套系统的交通大数据体系框架,如图1所示。

交通大数据研究,为提升现有数据的应用开发水平和效率带来了新的机遇,而大数据技术本身,也对进一步提升交通管理效率和公众服务水平提供了更加厚重的支撑。

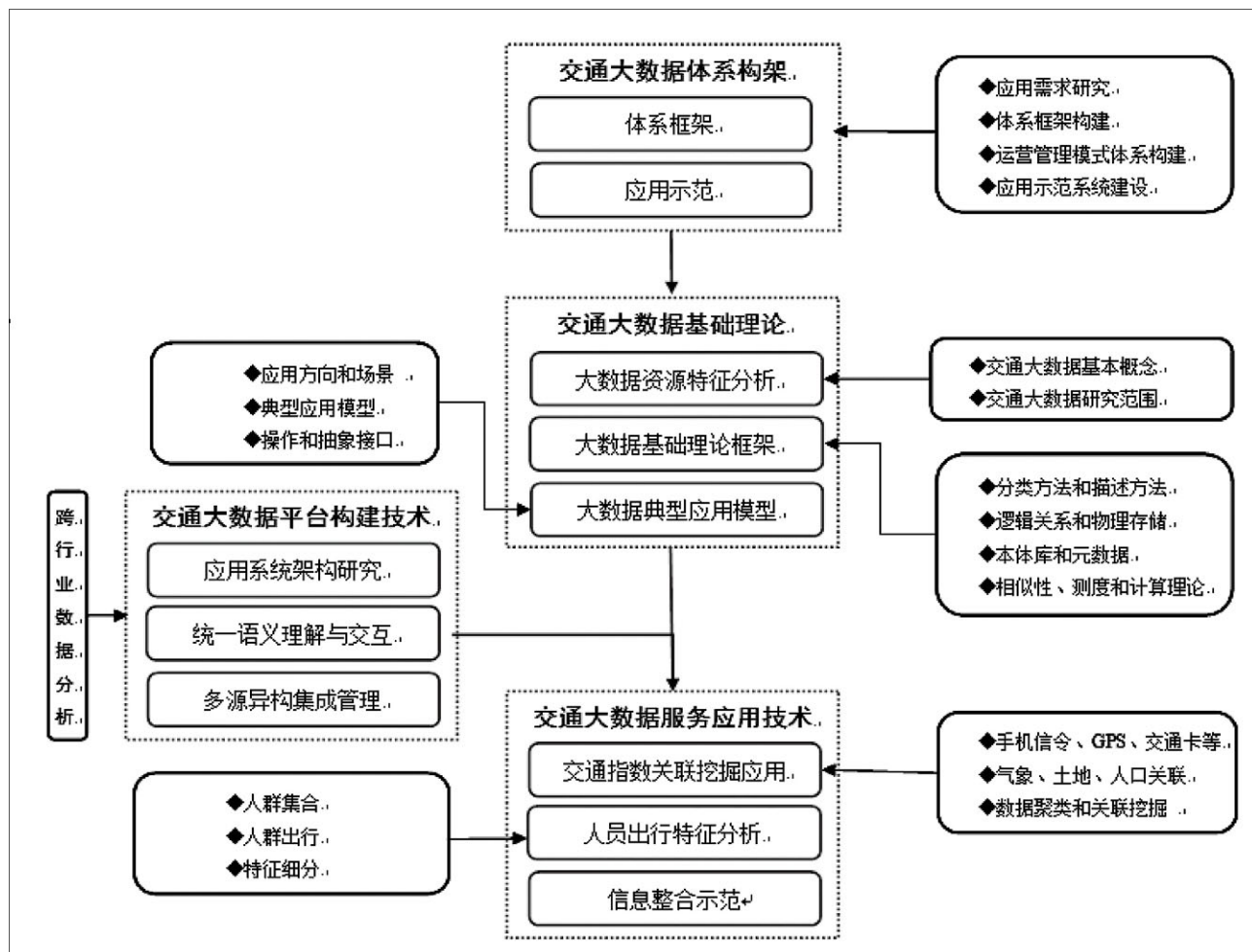


图1 交通大数据体系框架

3 交通大数据研究及示范应用

在构建的交通大数据体系框架下，本文对交通大数据理论基础、软硬件载体、跨领域数据关联挖掘等进行了一系列研究，主要得到了交通大数据本体库、基于Hadoop和Spark分布式内存的交通大数据应用示范平台、动态客流密度云图展示系统、《城市交通大数据》专著等一系列创新成果。

3.1 交通大数据本体库框架

交通大数据本体（或称为本体库）是对城市交通领域中共享概念和知识的抽象化、形式化的规范表示和说明^[6]。依据交通大数据的定义，交通大数据本体库框架包含三大部分，即：城市交通运行管理直接产生的数据，命名为“城市交通运营”类；城市交通相关的行业和领域导入的数据，命名为“相关领域导入”类；以及来自公众互动提供的交通状况数据，命名为“交通信息众筹”类。

交通大数据本体库涵盖城市交通规划、需求、综合以及城市道路交通、公共交通、对外交通等所涉及到的共3300多个概念。交通大数据本体库不仅可以为交通大

数据语义网络研究奠定基础，还能够帮助领域专家建立相应的交通行业新型管理系统与方法，提高交通信息自动化水平，为交通信息化研究、交通信息本体服务提供可以借鉴的框架和成果。

3.2 基于Hadoop和Spark的交通大数据应用示范平台

依托课题示范要求，建设了基于Hadoop和Spark分布式内存的交通大数据应用示范平台，提出并验证了基于Hadoop、关系数据库、内存库“多元合一”的交通大数据平台框架，实现了对结构化、半结构化、非结构化数据的集成管理与应用。采用Oracle、Spark分布式内存和Hadoop技术建成的交通大数据应用示范平台，其内存容量达到1.5TB、硬盘容量100TB，能够支持30TB以上的大数据计算。

基于同类数据等规模测试，对比分析了传统Oracle数据库环境与单纯Hadoop环境的数据处理能力，发现数据规模越大，大数据平台性能优势越显著，在300GB、15亿条车牌抓拍记录（相当于浦东卡口一年的数据量）数据测试中，示范平台性能显著优于对比环境。在车牌模糊匹配查询中，大数据示范平台能够达到独立

Hadoop 系统的 1/10 和 Oracle 数据库的 1/20 的速度。基于 Hadoop 的交通大数据示范平台强大的处理能力,为后续开展复杂交通数据关联分析和对比分析提供了基础。此外,系统的集成化数据管理环境,能够实现对示范系统数据的可视化组织与管理,为系统管理人员提供友好的工作交互环境。

3.3 交通大数据关联分析

交通大数据关联分析,目的在于发现城市交通与其它事物之间的联系,挖掘潜在未知的规律性,提高多源多类数据的利用率和价值,为交通管理部门提供更全面的数据支持与辅助决策。

采用回归分析与 Apriori 关联规则算法,对湿度、风速、雨量、温度等气象数据与交通状态指数进行关联分析,挖掘出了雨雪晴等不同天气与交通拥堵状态之间的量化关系,可以在可能的特大暴雨、暴雪等恶劣天气环境前,对重点路段或常发性拥堵区域提供交通疏导建议,也为依据气象预报对交通指数进行预测提供基础。通过采集志愿者的手机大数据,结合问卷调查,建立了基于人工神经网络、支持向量机、决策树等多个面向出行方式识别的数据挖掘模型,不仅能够识别轨道交通、小汽车、慢行交通等七类出行方式,还可以辅助交通部门完成城市居民出行方式等交通调查。以边界为 1 km 的正方形网格划分上海全市域,通过分析海量手机数据,计算出各网格内的客流密度和进出客流量,结合大数据可视化技术,开发出客流动态云图展示系统,能够实时把握各区域客流密度和客流量,找出城市人口职住分布、客流迁移与道路交通拥堵之间的关系,并对可能产生的突发大客流预警等。

3.4 交通大数据可视化系统

交通大数据可视化系统,以上海交通状态指数为应用案例,通过构建数据仓库,对指数本身及其与天气、事故、车流量等数据的关联分析可视化,从全网和指数区域两个层面,找出交通拥堵随时间和空间的变化态势。基于地理信息系统 GIS 和数据仓库技术的可视化系统,能够快速有效地浏览、查询、分析、挖掘交通指数数据,并易于人机交互,有效发现隐藏在交通大数据内部的特征和规律,从而支撑决策分析。

交通大数据可视化系统,是数据查询、分析、展示的综合体,是探索大数据信息转化为视觉模式,提高图表快速理解能力的研究成果之一。系统将各类交通数据作为指标,通过可视化拖拽,从不同维度对每项指标进行观察和分析,从而可以更加灵活地研究数据潜在规律。

3.5 《城市交通大数据》专著

基于这两年交通大数据的研究过程和成果,《城市交通大数据》得以编著并出版。该书从交通大数据的基本概念和研究应用现状出发,分析了大数据时代下的城市

交通需求、交通大数据中各类数据资源,介绍了一系列适合交通大数据组织、描述、管理、处理、分析挖掘和可视化等技术与方法,并给出了交通大数据应用开发的系统框架、方案选型、平台设计等各类参考和建议,提出了若干个具有代表性的典型交通大数据案例。全书共九章,50 余万字。

4 交通大数据应用展望

交通大数据的探索、研究和发展,直到最后形成实践和应用,不应该仅限于在一座城市、一个地区、一个领域或行业内部,应该从更高的视角出发来理解。基于上海交通大数据的研究和成果,本文认为,交通大数据的应用方向,应该与交通信息服务和产业化、智慧城市建设密切相关。只有这样,才能充分彰显和发挥交通大数据真正的活力^[7]。

4.1 交通大数据与交通信息服务产业化

交通大数据应用的一个重要方向是支撑交通信息服务的产业化发展。以开放、合作为原则,面向政府、企业、公众不同层面提供具有公益性或增值价值的数据和应用服务的新模式,形成“交通大数据魔方”,支撑交通大数据二次开发和产业化发展。这会涵盖 3 个方面的创新,即,数据集市服务模式创新、应用集市服务模式创新,以及交通大数据平台运营模式创新。

交通大数据运营模式,指政府与掌握全国性交通信息服务数据的行业领先企业共同进行运营管理的创新模式。期待形成面向社会提供公益性全方位的公众信息服务、面向企业提供增值的数据资源,以及分享数据融合成果的开放平台,也就是所谓的“大数据魔方”。

4.2 交通大数据与智慧城市

交通信息化的发展借助于信息技术、交通大数据的广泛应用,将实现以信息化为纽带的人(出行者)、车(出行设备)、路(出行设施)协同目标,将形成庞大的产业链,成为国家战略性新兴产业。利用交通大数据,可以支持交通政策和管理的精细化决策,辅助交通设施规划和建设实现更科学、优化的设计,实现基于智能车辆和道路技术的智能化交通协同,提高各种客货运服务和公众出行全过程交通信息的便捷、安全和个性化的服务。

作为智慧城市的重要一环,交通大数据的建设在导入其他行业数据的同时也为其他行业提供交通数据服务,打通城市运行的核心环节,从而全面地支撑智慧城市的建设与发展。

参考文献:

- [1] 上海市科学技术委员会. 上海推进大数据研究与发展三年行动计划(2013-2015 年)[R]. 2013.

交通信号倒计时应用研究综述

董利军, 吴稼豪, 马江山
(上海海事大学, 上海 201306)

摘要: 当下, 交通信号倒计时的使用越来越多。但是, 国内外学术界和管理部门对于倒计时装置的优缺点以及是否应该采用, 这些问题, 一直存在大量争论。从交通安全和交通效率两方面进行文献综述与分析, 总结已有的研究对机动车倒计时信号控制运行优劣的研究, 提出关于倒计时装置的使用建议, 以及今后可探讨的问题。

关键词: 机动车倒计时; 交通安全; 交通效率; 交叉口

中图分类号: U491 **文献标识码:** A **文章编号:** 1671-3400(2015)12-0053-04

A Literature Review of the Vehicle Traffic Signal Count-Down Applications

DONG Li-jun, WU Jia-hao, MA Jiang-shan
(Shanghai Maritime University, Shanghai 201306, China)

Abstract: The traffic signal count-down device has become more and more popular in China. However, there is a lot of debates about the advantages and disadvantages of the vehicle signal count-down device and whether it should be installed based on the traffic safety and operation efficiency point of view. In this paper, the merits of the count-down in traffic safety and traffic efficiency is summarized, and some suggestions on the use of this device are put forward.

Keyword: Vehicle signal Count-down; Traffic safety; Traffic efficiency; Intersection

0 引言

交通信号倒计时, 是指在普通交通信号灯旁设置数字或模拟光带倒计时的装置, 随着信号机的运行同步显

示现有信号相位(红灯或绿灯)的剩余时间。我国由于许多交叉口的信号周期过长, 常设在 220 ~ 240 s, 因而采用交通信号倒计时来指挥行人与车辆正常行驶, 提醒当下信号相位所剩的时间, 帮助驾驶员提前做出起步或者停车的决策。常用信号倒计时装置包括绿灯倒计时、红灯倒计时和早启绿闪等类型。

虽然国内外学者对于倒计时进行了大量的研究, 但是, 并没有对这些研究进行系统性的对比与总结。本文

收稿日期: 2015-08-24

作者简介: 董利军(1992-), 男, 甘肃白银人, 上海海事大学交通运输学院, 硕士在读, 主要研究方向: 交通控制研究。

- [2] 上海交通出行网. 畅通交通, 和谐上海 [EB/OL]. [2015-10-10]. <http://www.jtcx.sh.cn/>
- [3] 世博会交通协调保障组. 上海世博交通 [R]. 2011.
- [4] 薛美根, 朱洪, 邵丹. 上海世博交通研判技术与实践 [M]. 上海: 上海社会科学出版社, 2012.
- [5] 香山科学会议. 数据科学与大数据的科学原理及发展前景——香山科学会议第 462 次学术讨论会综述 [EB/OL]. [2015-10-10]. <http://www.xssc.ac.cn/ReadBrief.aspx?ItemID=1060>.
- [6] Wikipedia. Ontology (information science) [EB/OL]. [2015-10-10]. [http://en.wikipedia.org/wiki/Ontology_\(information_science\)](http://en.wikipedia.org/wiki/Ontology_(information_science)).
- [7] 何承, 朱扬勇. 城市交通大数据 [M]. 上海: 上海科学技术出版社, 2015.