

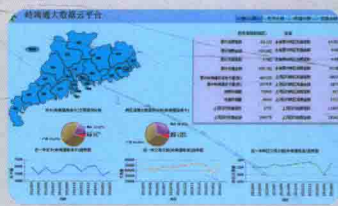
城市智能交通设计与实践技术丛书

CHENGSHI JIAOTONG YIKATONG

DASHUJU YINGYONG

城市交通

一卡通大数据应用



谢振东 李之明 徐锋 吴金成 刘雪琴 编著



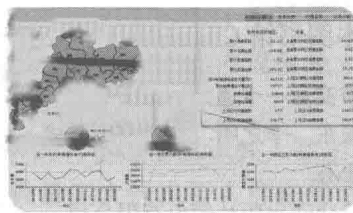
人民交通出版社股份有限公司
China Communications Press Co., Ltd.

城市智能交通设

CHENGSHI JIAOTONG YIKATONG
DASHUJU YINGYONG

城市交通

一卡通大数据应用



谢振东 李之明 徐锋 吴金成 刘雪琴 编著



人民交通出版社股份有限公司
China Communications Press Co., Ltd.

内 容 提 要

大数据的发展与应用逐渐成为人们的普遍共识,未来城市交通一卡通大数据的应用将是在大数据技术浪潮的推动下城市交通一卡通发展到一定阶段的必然产物,并且可以预见,城市交通一卡通大数据将在城市智能交通发展、一卡通行业发展、一卡通产业链发展、公众智慧出行等方面发挥重要作用。

本书从概念、特征、需求、技术、应用等多方面剖析和延伸城市交通一卡通大数据的内在价值,重点突出交通一卡通大数据在提高城市交通的智能化水平、促进一卡通产业数据资源的整合、提高公众出行的便利度、实现产业链的优化升级等方面所起的关键作用和重要意义,为一卡通行业大数据的发展提供参考和借鉴。

图书在版编目(CIP)数据

城市交通一卡通大数据应用 / 谢振东等编著. —北京:人民交通出版社股份有限公司, 2016.5

ISBN 978-7-114-13015-1

I. ①城… II. ①谢… III. ①城市交通—数据处理—研究 IV. ①U12

中国版本图书馆 CIP 数据核字(2016)第 103306 号

书 名: 城市交通一卡通大数据应用

著 作 者: 谢振东 李之明 徐锋 吴金成 刘雪琴

责任编辑: 刘永芬

出版发行: 人民交通出版社股份有限公司

地 址: (100011)北京市朝阳区安定门外外馆斜街 3 号

网 址: <http://www.ccpres.com.cn>

销售电话: (010)59757973

总 经 销: 人民交通出版社股份有限公司发行部

经 销: 各地新华书店

印 刷: 北京市密东印刷有限公司

开 本: 720×960 1/16

印 张: 19.75

字 数: 350 千

版 次: 2016 年 6 月 第 1 版

印 次: 2016 年 6 月 第 1 次印刷

书 号: ISBN 978-7-114-13015-1

定 价: 48.00 元

(有印刷、装订质量问题的图书由本公司负责调换)

前 言

城市交通一卡通在我国起步于1999年,最近几年发展十分迅速。目前全国累计有超过440个城市建立了不同规模的公交IC系统,已有254个地市级以上城市发行了城市交通一卡通,全国累计发卡量超过7.5亿张,其中涉及跨区域实现互联互通的公交卡累计发行1.5亿张,使用范围覆盖了公共交通领域的公交、地铁、轮渡、出租汽车、有轨电车、城市公共自行车等多种交通方式,应用功能扩展到小额消费领域(如餐饮、便利店、休闲娱乐等)、公共管理领域(如企业、社区、医院、校园等范围内的门禁、缴费、考勤、身份识别等)以及电子商务领域(在线支付、在线充值、在线缴费等)。随着城市交通一卡通的发展,未来实现全国范围内的互联互通已经成为必然的趋势。作为全国领先的跨区域交通互联互通典范的广东省公交一卡通——岭南通已覆盖广州、深圳、东莞、佛山、肇庆等省内21个地市,并与香港、澳门实现互联互通,累计发卡量超过4700万张,成为中国国内规模最大的区域交通一卡通系统。

在当前大数据时代的背景下,越来越多的行业逐渐意识到大数据潜在的价值。城市交通一卡通领域也不例外,面对每天数量庞大的出行人群在使用城市交通一卡通过程中产生的大量交易数据,政府部门、企业机构、科研单位纷纷开始思考并研究这些数据蕴藏着的内在价值。随着信息技术的发展,城市交通信息化以及智慧城市建设的不断深入,各个城市交通运行管理所产生的数据、相关行业和领域形成的数据以及公众互动过程的数据均呈爆炸式的增长,这些数据将成为城市交通一卡通数据分析过程中不可缺少的关联数据。若这些数据通过不同渠道、网络汇集在一起,形成大数据的聚合效应,将是一笔巨大的财富和资源。目前,大数据的发展与应用已上升到国家战略的高度,引起了全社会各行各业的普遍关注,编者认为未来城市交通一卡通大数据的应用将是在大数据技术浪潮的推动下城市交通一卡通发展到一定阶段的必然产物,并且可以预见,城市交通一卡通大数据将在城市智能交通发展、一卡通行业发展、一卡通产业链发展、公众智慧出行等方面发挥重要作用。

基于以上背景的研究与分析,城市交通一卡通大数据的研究、分析与应用具有十分重要的意义,主要体现在:

1) 有助于促进城市智能交通系统的建设,提高城市交通的智能化水平

城市交通一卡通数据蕴含着丰富的城市公共交通客流信息、乘客的出行规律等信息,这些信息在智能交通系统有着非常重要的应用。美国、韩国、日本等国家,国内北京、上海、广州等城市都建立了智能交通系统,这些国家或城市的城市交通一卡通系统都具有较好的应用基础,并在智能交通系统的建设中实现了城市交通一卡通数据的应用,大大提高了城市交通的智能化水平。以广州为例,广州基于“羊城通”刷卡数据建立了公交线网分析系统,系统通过刷卡数据提取公交客流特征,并根据影响特征分布进行客流预测,通过客流分配模型完成客流在公交线网上的分布模拟,通过生成与优化模型完成线网优化调整,通过评价模型对规划结果做综合评价,并将最终结果在分析系统上实现模拟仿真。在未来,借助城市交通一卡通大数据,城市智能交通系统的建设水平必将更上一层楼,大大提高城市交通的智能化水平。

2) 有助于促进一卡通产业数据资源的整合,实现产业链的优化升级

城市交通一卡通在我国经过几十年的发展,得到了大力推广与应用,目前已经取得了很好的社会经济效益,并以此逐渐形成庞大的产业链。随着我国经济结构体制改革的逐步深化,各行业都在探索挖掘新的业务增长点以提升自身实力并在新的经济浪潮中占领先机,一卡通产业链上的各大企业也不例外。城市交通一卡通大数据可以对一卡通产业链上的供需信息、风险信息、企业信息等信息进行有效的整合,实现产业链各环节数据的互通,有助于省去原有产业链的过度环节,优化产业链结构,拉近产业链上下游的关系,打造效率更高、架构更优、沟通更顺的一卡通产业链。在此背景下,城市交通一卡通大数据势必成为一卡通产业链中重要的大数据资源,实现一卡通产业数据资源的整合,促进一卡通产业的产业升级与产业优化,进一步提高资源的利用率,实现一卡通产业健康、良好、有序的发展。

3) 有助于促进公众智慧出行系统的建设,提高公众出行的便利度

随着城市交通的发展和人们对交通出行要求的提高,传统的出行方式已经无法满足公众庞大出行的需求。随着智慧城市概念的推出,众多城市也不断加入到智慧城市的建设中。基于这个大背景,人们提出了智慧出行的创新理念,即希望建立一个智慧出行服务平台,可以为出行民众提供交通信息、实时路况信息、换乘信息、交通气象信息、停车场信息以及与出行相关的其他信息,使出行者可以根据这些实时信息灵活安排与规划自己的出行线路,而基于城市交通一卡通系统产生的数据正好满足这种需求。因此,建设基于城市交通一卡通大数据的智慧出行服务平台,将有效满足大数据量、大用户量的系统应用,为用户提供持续、有效、动态、便捷的出行信息服务,提高公众出行的效率。

本书从当下热门讨论的大数据技术与应用的角度,引入城市交通一卡通大数

据的应用与发展,沿着这个思路,本书结构安排如下:第1章主要介绍大数据的发展概况,进而引入大数据在交通领域的应用,重点从交通大数据概述、特征分类、采集方式及实践等方面论述;第2章进一步细化切入交通一卡通的应用主题,介绍了交通一卡通数据来源、特点、管理等;第3章探求了交通一卡通大数据的研究与应用,涉及内容包括一卡通大数据需求、国内外对一卡通数据研究与应用等;第4章论述了交通一卡通大数据技术体系,从技术框架、采集方式、数据预处理等方面构建交通一卡通的数据分析体系,并配以一些实际案例来说明大数据的应用;第5章从不同需求论述交通一卡通大数据服务;第6章归纳了城市交通一卡通大数据发展过程中的机遇与挑战,并在此基础上对城市交通一卡通大数据的发展进行展望和思维探索。

本书作为《城市智能交通设计与实践技术丛书》之一,与其他各册共同构成有机统一的整体,是对其他分册有益的补充。

本书适合城市交通一卡通大数据领域的相关技术人员阅读,同时也可作为城市交通一卡通相关行业从业人员决策、管理和研究的参考书。

本书的编著小组主要成员包括谢振东、李之明、徐锋、吴金成、刘雪琴。另外,方秋水、杨晓丽、常振廷、谭丹丹、曾烨、余红玲、刘强、何建兵、张景奎等人员亦对本书提供了宝贵的意见和技术指导,在此表示衷心的感谢,同时也感谢广州羊城通有限公司等岭南通产业联盟成员单位的大力支持和帮助。

由于编写时间紧迫、任务重,编者在城市交通一卡通大数据领域的研究还不够深入,因此,书中选材、论述、引用等可能存在不当或错误的地方,望广大读者能够多加理解,并及时联系编者以便修正,以期在后续出版中进行完善。

编者
2016年3月

目 录

第1章 交通大数据概述	1
1.1 交通大数据内涵	1
1.2 交通大数据特征与分类	5
1.3 交通大数据的采集方式	13
1.4 交通大数据的实践	19
第2章 大数据时代下的城市交通一卡通大数据	32
2.1 城市交通一卡通大数据定义及特点	32
2.2 城市交通一卡通大数据来源	33
2.3 城市交通一卡通大数据的管理	61
2.4 城市交通一卡通大数据平台构建	66
第3章 城市交通一卡通大数据研究与应用	92
3.1 需求分析	92
3.2 国内外研究	103
3.3 主要城市交通一卡通大数据应用	110
第4章 城市交通一卡通大数据技术体系	122
4.1 大数据技术框架	122
4.2 一卡通大数据采集	122
4.3 大数据预处理	127
4.4 大数据存储与管理	144
4.5 大数据分析挖掘	152
4.6 大数据可视化	195
第5章 城市一卡通大数据服务	210
5.1 城市交通规划与建设	210
5.2 城市交通管理	215
5.3 交通决策	217
5.4 一卡通产业链服务	221
5.5 一卡通应用服务领域分析	233
5.6 公众智慧出行服务	239

第 6 章 城市交通一卡通大数据展望	252
6.1 概述	252
6.2 机遇	252
6.3 挑战	253
6.4 展望	258
6.5 探索	263
第 7 章 城市交通大数据大赛与应用案例分享	268
7.1 广东公共交通大数据竞赛背景	268
7.2 大数据竞赛简介	269
7.3 大数据竞赛优秀算法分享	274
附录 A 术语说明	282
附录 B 缩略语	287
附录 C IC 卡相关标准	290
附录 D 全球十大著名的数据中心	291
参考文献	297

第1章 交通大数据概述

随着信息技术的迅猛发展和普及应用,移动互联网、云计算、物联网、社交网络、智能终端等技术和产品相继融入我们日常的工作和生活,在这个过程中所产生的视频、音频、图像、日志等数据信息量呈现指数式爆炸增长之势。国际数据公司(IDC)的研究结果表明,2008年全球产生的数据量为0.49ZB,2009年的数据量为0.8ZB,2010年增长为1.2ZB,2011年的数据量更是高达1.82ZB,据推算,到2020年底,全世界所产生的数据规模将达到35.2ZB。

丰富的数据来源和爆炸式的数据增长促使大数据出现,并得到广泛的关注和应用。大数据以其多源、海量、异构等特性冲击着社会的各个领域,为传统的数据库系统在存储、访问、管理、挖掘分析大数据方面带来巨大的考验,同时,大数据的分析与利用突破了信息技术产业的局限性,在政府、医疗、互联网、制造业、金融业、运输业等诸多领域和行业得到了广泛的应用。

大数据以前所未有的速度,带来了思维变革、管理变革及商业变革,驱动了产业间的融合与分立,颠覆了人们探索世界的方法,然而大数据时代的到来并不是突如其来的,它在人类进步的大潮中酝酿而生。本章将介绍大数据的基础知识,使读者对大数据的起源、发展历程、定义及特征、发展现状与趋势等内容有基础的认识和了解,从而对大数据的应用与发展产生兴趣。

城市化的快速发展,特别是大城市人口的扩张,使得本来有限的资源变得更为紧张,这种情况在交通领域表现得更加突出。城市交通拥堵、交通污染日益严重、交通事故频繁发生,这些都是各大城市亟待解决的问题。智能交通成为改善城市交通的关键所在。为此,及时、准确获取交通数据并构建交通数据处理模型是建设智能交通的前提,通过大数据技术的发展与应用可使上述问题得以有效解决。

1.1 交通大数据内涵

1.1.1 大数据概念的演化

在论述交通大数据之前,应先了解什么是数据。一般来说,数据是指存储在某种媒体或介质上能够被识别的物理符号,包括数据的内容和数据的形式两个方面。数据的内容是所描述的客观事物的具体特质,数据的形式是指数据存储在某种媒体上的具体形式。数据不仅包含数字、字母、文字等文本展示形式,而且还包

括多媒体数据,如图形、图像、动画、影像、声音等。“大数据”一词由英文单词“Big Data”翻译而来,“Big”是形容词,意为大的、庞大的、巨大的、大量的,“Data”为名词,意为资料、数据。在大数据研究过程中,也有学者将“Big Data”翻译为“海量数据”或“巨量数据”,主要突出了其数量上的规模,但不足以概括“Big Data”的丰富内涵,相反,“大数据”这一翻译简单直白同时也更贴近“Big Data”实质内涵的广阔性、复杂性、抽象性和神秘性,故而在学术界的认可度较高。

虽然数据在日常工作和生活中无处不在,也经常能接触到,然而对于大数据的概念,却少有人能说得清楚。那么,是谁最早提出大数据这一概念,或者说大数据这一概念是如何产生的,从数据发展到大数据经历了哪些过程?

大数据的概念进化史是一个漫长的过程,有一点可以肯定:大数据是由一般原始数据进化过来的。伴随着科学技术的发展与进步,特别是移动互联网、云计算等技术的日益普及和应用,全球数据总量不断扩展,增长速度日渐加速,人们开始思考一个问题:数据量积累到某个程度是否会产生某些预料不及的后果或价值,人们甚至开始担忧有一天会出现“数据爆炸”,这种担心并不是没有道理的。

卫斯理大学图书馆员弗莱蒙特·雷德在其1944年出版的《学者与研究型图书馆的未来》书中估计:美国高校图书馆的规模每16年就会翻一番。按照这一增长速度,雷德推测2040年耶鲁大学图书馆将拥有“约2亿册藏书,将占据6000余英里书架,需要的编目人员超过6000人”,这无疑将对世界图书的存储和管理产生巨大的挑战。

德里克·普赖斯在其1961年出版的《巴比伦以来的科学》书中,通过观察科学期刊和论文的增长规律来研究科学知识的增长。他得出以下结论:新期刊的数量以指数方式增长而不是以线性方式增长,每15年翻一番,每50年以10为指数倍增长。

由此可见,20世纪的学者已经意识到数据增长的速度之快与未来数据量的庞大,并对这种增长速度表示担忧。

在经过一系列的探索与讨论后,人类社会仍没有找到解决数据量增长带来的种种问题的办法,依然对存储和利用这些数据感到迷茫。这一阶段为大数据的萌芽阶段。到了1980年,著名未来学家阿尔文·托夫勒在《第三次浪潮》一书中,将大数据热情地赞颂为“第三次浪潮的华彩乐章”。阿尔文·托夫勒在这本书中使用了大数据这个词,但此时对大数据的理解比较模糊,大数据与大量数据存在着模糊的界限,没人能够说明大数据与大量数据有什么区别。作者更多只是将“大数据”简单地意为“量大的数据”。在这之后,学者继续对数据的增长、大量数据的存储进行研究分析。

2001年2月,梅塔集团分析师道格·莱尼发布了一份研究报告,题为《3D数据

管理;控制数据容量、处理速度及数据种类》。这份报告提出了3D数据的3V特征,即数据容量大、处理速度快和种类多。10年后,3V作为定义大数据的三个维度而被广泛接受。

2010年2月,肯尼斯·库克尔在《经济学人》上发表了一份关于管理信息的特别报告《数据,无所不在的数据》。库克尔在文中写道:“……世界上有着无法想象的巨量数字信息,并以极快的速度增长……从经济界到科学界,从政府部门到艺术领域,很多地方都已感受到了这种巨量信息的影响。科学家和计算机工程师已经为这个现象创造了一个新词汇——‘大数据’。”

2011年5月,全球知名咨询公司麦肯锡发布了一份题为《Big data: The next frontier for innovation, competition and productivity》(大数据:下一个竞争、创新和生产力的前沿领域)的研究报告。在该报告中,麦肯锡给出了大数据的定义:大数据是指无法在一定时间内用传统数据库软件工具对其内容进行采集、存储、管理和分析的数据集合。同时指出,数据已经渗透到每一个行业和业务职能领域,逐渐成为重要的生产因素;而人们对于海量数据的运用将预示着新一波生产率增长和消费者盈余浪潮的到来。这份报告最早使用今天为大家理解的“大数据”概念并提出了大数据时代已经到来。报告发布后,大数据迅速成为计算机行业争相传诵的热门概念,也引起了其他各界的高度关注。

自此,大数据受到了越来越多的研究与关注,人们渐渐明白大数据与大量数据之间的差异,大数据的概念、特征等也渐渐明晰,大数据受到世界的重视。

1.1.2 交通大数据概念

所谓交通大数据,简单来说就是通过各种途径收集和整合交通出行、运营及行政管理数据,通过大数据分析技术,使其应用于交通管理的各个领域,形成一套具有交通领域特色的数据分析体系和应用主题,为完善交通部门管理、企业运营、公众出行服务提供数据支撑。

社会经济的快速发展促使城市机动车辆的数量大幅增加,城镇化的加速打破了城市道路系统的均衡状态,传统的交通管理信息系统难以满足当前复杂的交通需求,交通堵塞成为棘手问题。而长尾理论告诉我们,大数据将海量数据聚合在一起,将离散的数据需求聚合能形成数据长尾,从而满足传统中难以实现的需求。例如交通需求,个体对某一道路路段的特定信息需求,尽管人数不多、份额较少,但聚合效应加长了数据的长尾,这些个体的需求数据会在数据的长尾下形成庞大的需求市场,为满足人群的需求以及个性化服务奠定了数据基础。可见,用大数据管理交通是交通管理模式的变革,与此同时也变革了公共交通市场管理的整个内涵,未来交通管理呼唤大数据的应用。

交通大数据应用的本质是在原有的道路交通资源的基础上,通过挖掘数据的潜在价值,分析交通行为规律,优化交通管理结构,改善交通环境,提升交通智能化水平,为人们提供更为人性化的出行服务,降低交通出行成本。

推动交通大数据实施的关键基础包括三个方面:设备的连接、大数据管理以及数据开放。

在“大数据”时代,人与人、人与设备以及设备之间都依靠数字网络进行连接,通过这些连接不断产生数据,是推动社会管理创新、经济结构优化的重要动力。在交通领域,公共交通数据的获取途径有无线网络、传感设备、有线采集等方式。例如在车联网的背景下,通过车辆连通互联网来上传数据或下载数据,或是车辆与车辆之间进行数据共享。这些都有赖于整个地区网络速度的提升,以提高数据的传送速度,减少所有车辆的宽带费用。还有一种可能的解决方案是,在每辆车内置WiFi以便车辆之间共享数据,每辆车也可以从电视广播中获取数据。

大数据管理是一个系统工程,是基于一种操纵和管理大数据的数据库大型软件,用于建立、使用和维护大数据仓库,对数据库进行统一的管理和控制,以保证数据库的安全性和完整性,它可使多个应用程序和用户用不同的方法在同时或不同时刻去访问数据库。但仅依靠技术是难以解决问题并带来高效益的。需要通过高效的管理,使该管理系统可连续、实时地采集相关数据,将这些数据整合在大数据管理系统中,以便进行科学分析,并及时将智能分析后的用户个性化需求结果转交用户,使用户、交通线路以及拥堵影响因素有机聚合为相互关联的、物联化与智能化的交通运输体系。

维基百科指出“开放数据具有这样一种思想:对任何人来说,他们都能如自己所愿那样免费使用和再次发布某些数据,而不用受版权、专利或其他控制机制的限制”。开放数据所提倡的是一种积极的信息公开理念,可公开的数据必须公开,以使其发挥更大的作用。只有将与交通相关的所有信息进行公开,才有助于大数据管理系统科学地分析问题,提供被分析的数据越多越及时,所得出的结论就可能越精确。若没有与交通相关的大量公共信息的公开,大数据管理将是无源之水。

城市交通信息数据系统是基于大数据应用技术的交通行业信息共享交换中心,数据中心建立以后,将成为城市交通信息的枢纽。

城市交通综合监测和预警系统可以实现对整个城市交通状况的实时监测。交通管理部门可以对城市交通中可能发生的大面积交通瘫痪做出有效的预判。同时,该系统也可以引导公众出行,为公众提供全面、及时的出行信息,真正达到绿色交通的出行要求。

交通大数据包括道路信息数据、GPS 导航数据、客运数据信息、公交运营信息、站场数据、客流信息、一卡通数据、交通监控数据、车辆数据、从业人员数据、诚信从

业信息数据、交通工程数据、交通执法数据等,这些数据分布在不同的部门,需要通过一个有效机制进行融合,打破数据孤岛,形成具有新的应用价值的大数据资源宝库(图 1-1)。

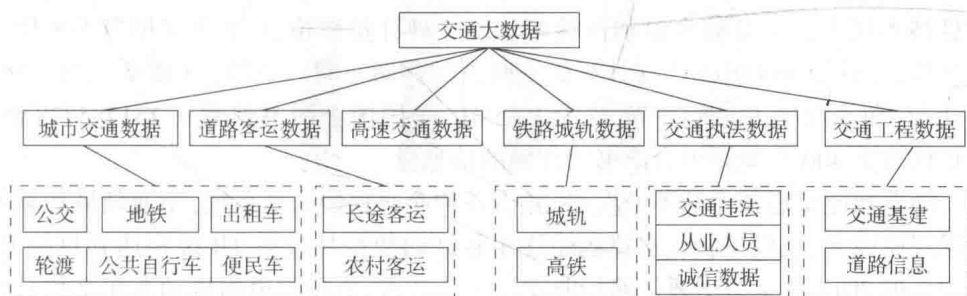


图 1-1 交通大数据框架

1.2 交通大数据特征与分类

大数据具有四大特征,分别是规模大(Volume)、数据类型多样(Variety)、速度快(Velocity)、价值高(Value),即所谓的 4V 特性(图 1-2)。虽然后续不断有人补充对大数据特征的理解,如准确(Veracity),强调处理的数据必须准确而精确,这样不仅可以保证数据的处理速度和时间,还可以保证分析决策的正确性;也如多样性(Variability),指出数据流的不一致对数据加载造成挑战,特定事件的发生会导致数据量急剧增加而使系统维护变得困难,比如淘宝“双十一”产生的数据量对系统的承载力就是很大的考验等;也有学者给大数据增加了“C”特性,即复杂(Complexity),认为大数据的结构相互关联和相互依存的程度不一,一个或几个元素的微小变化可能会产生非常大的扰动。虽然来自各行业研究人员对大数据特征的诠释和补充都有各自的道理和适用的合理性,但总的来说都不及“4V”具有代表性。

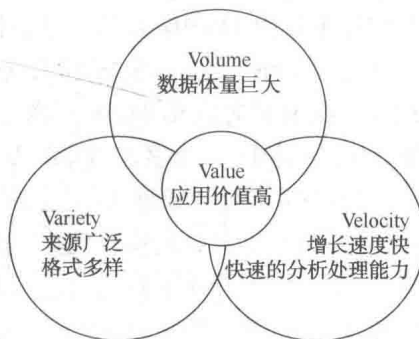


图 1-2 大数据的 4V 特征

1.2.1 交通大数据特征

经上节分析大数据的特征后,对大数据的内涵与概念的延展也有了一定的认识,不妨将大数据的触角延伸至交通领域,具体分析交通大数据的基本特征和表现。

1) 规模大 (Volume)

在说大数据规模大的特性之前,不得不提一下数据的计量单位:字节。现有数据大多存储于计算机中,计算机是采用二进制数进行工作的,字节(Byte)是计算机信息技术用于计量存储容量和传输容量的一种计量单位,1个字节相当于8位二进制数,也就是“00001001”在计算机中所占的存储空间。一部《红楼梦》含标点约87万字,每个汉字占2个字节,约为1.65MB,美国国会图书馆藏书151785778册,1EB相当于400个美国国会图书馆存储的信息量。

随着智能交通技术的不断发展,凭借各种交通数据采集系统,交通领域积累的数据规模庞大,飞机、列车、水路运输逐年累计的数据从过去TB级别达到目前PB级别。据2014年全国交通运输行业发展统计公布,全国国道网机动车年平均日交通量为15157辆(当量标准小客车,下同),按可比口径(下同)比2013年增长3.8%。全国国道网机动车日平均行驶量为186538万公里(当量标准小客车,下同),增长2.4%。其中,国家高速公路日平均交通量为23551辆,日平均行驶量为97423万公里,年平均交通拥挤度为0.39,分别增长5.5%、4.2%和5.7%;普通国道日平均交通量为10899辆,日平均行驶量为89048万公里,年平均交通拥挤度为0.63,分别增长1.9%、0.5%和持平。全国高速公路日平均交通量为22021辆,日平均行驶量为119894万公里,年平均交通拥挤度为0.36,比2013年分别增长7.1%、6.1%和7.4%。全国拥有公路营运汽车1537.93万辆,比2013年末增长2.2%。拥有载客汽车84.58万辆、2189.55万客位,比2013年末分别减少0.8%和增加0.9%。其中,大型客车30.67万辆、1326.24万客位,分别增长2.6%和3.4%。

2014年末统计数据表明,全国城市及县城拥有公共汽电车52.88万辆、59.79万标台,比上年末分别增长3.8%和4.3%,其中BRT车辆5339辆,增长19.1%。全国有22个城市开通了轨道交通,2014年新开通3个。拥有轨道交通车站1829个,增加280个,其中换乘站151个,增加17个;运营车辆17300辆、41770标台,其中,地铁车辆15696辆,轻轨车辆1372辆。出租汽车运营车辆137.01万辆,城市客运轮渡329艘。全年全国铁路完成旅客发送量23.57亿人,如图1-3所示,旅客周转量11604.75亿人公里,比2013年分别增长11.9%和9.5%。其中,国家铁路完成23.24亿人,11556.36亿人公里,分别增长12.0%和9.5%。

这些庞大的数字背后隐藏一个巨大的数据黑洞,吸纳了海量数据,并在以几何级别的速度高速增长,若将这些数据进行分类整理和存储,通过大数据的分析,未来将产生不可估量的价值。

2014年广东全省公路客运量和周转量分别完成了15.7亿人和1629.8亿人公里,同比增长9.64%和11.41%;公路货运量和周转量分别完成了25.7亿吨和3113.8亿吨公里,同比增加7.38%和16.4%(来源:广东省交通运输厅数据统计)。

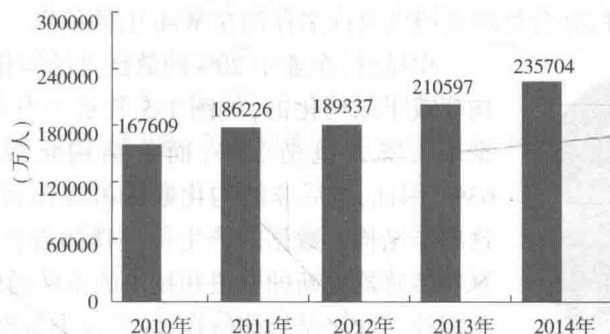


图 1-3 2010—2014 年铁路旅客发送量(来源:全国交通运输行业发展统计公报)

全国有 31 个城市的汽车数量超过 100 万辆,其中北京、天津、成都、深圳、上海、广州、苏州、杭州 8 个城市汽车数量超过 200 万辆,北京市汽车超过 500 万辆。2013 年,广州市内公交车总数达 10850 台,全市出租汽车总运力达到 21889 台。

无论从国家还是省市交通运输发展的现状来看,每天在交通领域中产生的数据是不可估量的,这将为交通大数据分析提供海量的数据来源,如图 1-4 所示。



图 1-4 2014—2020 年中国交通领域大数据市场规模及增长情况
(来源:贵阳大数据交易所)

2) 数据类型多样 (Variety)

造成数据类型多样的主要原因是数据来源广泛和数据格式多,不同的应用系统、不同的设备产生的数据类型各不相同,数据类型包括表格数据、分层数据、文件、电子邮件、计量数据、视频数据、音频数据、图片、金融交易和其他更多种类,这些多样化的数据从结构上大体可以分为三类:一是结构化数据,即行数据,存储在数据库里,可以用二维表结构来实现的数据,如一卡通交易数据、高速公路收费及通行数据、站场发班数据、交通从业人员数据等;二是非结构化的数据,如视频、图片、音频、网页等,如交通监控视频数据、交通违法拍照数据等;三是半结构化数

据,包括电子邮件、办公处理文档以及许多存储在 Web 上的信息。

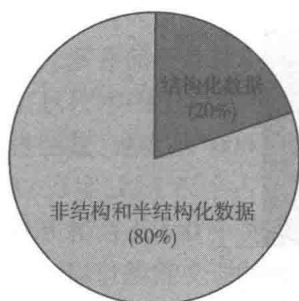


图 1-5 结构化数据

据统计,企业中 20%的数据是结构化的,80%是非结构化或半结构化的,如图 1-5 所示。当今世界结构化数据增长率大概是 32%,而非结构化数据增长率则是 63%,因此,今后非结构化数据占有比例还将继续增加,这些非结构化数据的产生往往伴随着社交网络、移动计算和传感器等新的渠道和技术的不断涌现和应用。现在利用较多的是结构化数据,因为这类数据规律性较强,数据整齐易于分析,但是大数据需要处理的数据不单单是结构化的数据,那样会导致分析结果不准确全面,如此多

样的数据对大数据分析处理带来很大的挑战。

3) 速度快 (Velocity)

该特征一方面体现了数据量呈现快速增长的趋势,更重要的是突出对数据要求具有快速的分析和处理能力。在网络时代,通过基于实现软件性能优化的高速电脑处理器和服务器的,创建实时数据流已成为流行趋势。政府、企业等机构不仅需要了解如何快速创建数据,还必须知道如何快速处理、分析数据,以满足自身的实际需求。大数据的动态性强,有些数据要求实时处理,例如发生火灾、地震等紧急事件,有些数据为决策支持,例如股票分析、导航定位等,这些都要求分析处理速度必须足够快,否则分析处理得到的结果可能就过时了,也就失去了对大数据进行分析原本应有的意义和价值。

据 2014 年交通运输行业发展统计公报显示,客运量和货运量发展迅速:全年全国营业性客运车辆完成公路客运量 190.82 亿人、旅客周转量 12084.10 亿人公里,比上年分别增长 3.0% 和 7.4%,平均运距 63.33 公里。全国营业性货运车辆完成货运量 333.28 亿吨、货物周转量 61016.62 亿吨公里,比上年分别增长 8.3% 和 9.5%,平均运距 183.08 公里。

全国交通运营线路和里程快速上升:全国拥有公共汽电车运营线路 45052 条,运营线路总长度 81.78 万公里,比上年末增加 3314 条、6.88 万公里,其中公交专用车道 6897.3 公里,增加 1006.7 公里;BRT 线路长度 2790.3 公里,增加 37.7 公里;轨道交通运营线路 92 条,运营线路总长度 2816.1 公里,增加 11 条、408.2 公里,其中地铁、轻轨线路分别为 76 条、2418.0 公里和 9 条、303.5 公里。城市客运轮渡运营航线 126 条,运营航线总长度 497.6 公里,比上年末分别减少 17 条、77.8 公里。全国公路客运量如图 1-6 所示。

城市客运需求快速增长:全年城市客运系统运送旅客 1315.66 亿人,比上年增长 2.5%。其中,公共汽电车完成 781.88 亿人,增长 1.4%;BRT 客运量 14.76 亿人

次,增长 34.7%;公共汽电车运营里程 346.69 亿公里,下降 0.6%;轨道交通完成 126.66 亿人,运营里程 3.27 亿公里,分别增长 16.0% 和 19.2%;出租汽车完成 406.06 亿人,运营里程 1618.11 亿公里,分别增长 1.0% 和 1.6%,平均每车次载客人数 1.96 人/车次,空驶率 31.2%;客运轮渡完成 1.07 亿人,增长 0.7%。

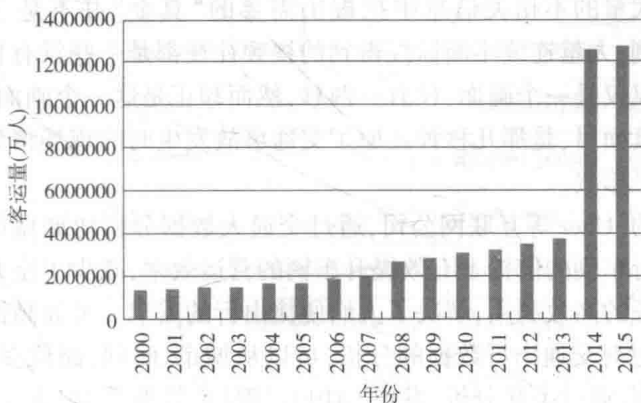


图 1-6 全国公路客运量

交通一卡通全国互联互通已提上日程:交通运输部发布的《关于促进交通一卡通健康发展加快实现互联互通的指导意见》中指出,推进以区域为重点的互联互通,选择京津冀、长三角、珠三角、长江经济带中游城市群等条件比较成熟、有内生需求的重点地区,以及公交都市创建城市,率先启动实施城市间交通一卡通互联互通工程。大力开展交通一卡通在出租汽车、长途客运、城际轨道、水上客运、公共自行车及停车场等交通运输领域的应用,积极推进在高速铁路、民航等领域的应用。到 2020 年基本实现各大城市群跨市域、跨省域的交通一卡通互联互通。未来全国交通一卡通互联互通将覆盖交通全出行链,形成无比庞大的交通出行数据,涉及数量之大、面积之广、人数之多将是前所未有的。

庞大的交通运输基数以及高速增长的交通基础设施建设、客运量、运营里程将大大推动交通大数据的增长,并对交通大数据分析技术提出更高的要求。

4) 价值高(Value)

价值是大数据的终极目的,大数据本身是一个“金矿”,可以从大数据的分析和处理中获得意想不到的有价值的信息,特别是在激烈竞争的商业领域,大数据正成为一种新型资产,各领域都在追求数据最大化。正因为如此,谁拥有数据,谁就有拥有资产。每个城市每天都产生海量的数据,而这些数据如果能够得到充分的应用,将有效提升城市公共交通的运行效率,减少不必要的道路拥挤,降低城市整体运行成本和人们的出行成本,体现交通大数据的应用价值。