

JIAOTONG YUNSHU

WULIU DASHUJU
FENXI YU YINGYONG

交通运输物流大数据 分析与应用

牟向伟 蒋晶晶 著



华中科技大学出版社
<http://www.hustp.com>

交通运输物流大数据分析与应用

牟向伟 蒋晶晶 著

华中科技大学出版社

华中科技大学出版社

中国·武汉

内 容 简 介

本书针对交通运输物流领域的大数据环境特点,对大数据存储以及大数据分析方法进行分析研究,设计了一种交通运输物流领域大数据基础技术架构,支持存储、计算、资源调度管理等各个层面的应用,以满足数据增值和软件开发的需要,最终将该技术架构实践于交通运输物流领域的流式数据实时计算分析与离线数据并行计算分析这两种典型应用中。

图书在版编目(CIP)数据

交通运输物流大数据分析与应用/牟向伟,蒋晶晶著. —武汉:华中科技大学出版社,2021.9
ISBN 978-7-5680-7520-6

I. ①交… II. ①牟… ②蒋… III. ①数据处理-应用-交通运输管理-研究 ②数据处理-应用-物流管理-研究 IV. ①F502 ②F252

中国版本图书馆 CIP 数据核字(2021)第 172124 号

交通运输物流大数据分析与应用

牟向伟 蒋晶晶 著

Jiaotong Yunshu Wuliu Dashuju Fenxi yu Yingyong

策划编辑: 聂亚文

责任编辑: 狄宝珠

封面设计: 孢子

责任监印: 徐露

出版发行: 华中科技大学出版社(中国·武汉)

电话: (027)81321913

武汉市东湖新技术开发区华工科技园

邮编: 430223

录排: 华中科技大学惠友文印中心

印刷: 武汉开心印印刷有限公司

开本: 710mm×1000mm 1/16

印张: 11

字数: 206 千字

版次: 2021 年 9 月第 1 版第 1 次印刷

定价: 48.00 元



本书若有印装质量问题, 请向出版社营销中心调换
全国免费服务热线: 400-6679-118 竭诚为您服务
版权所有 侵权必究

前 言

大数据背景下的相关问题已引起了信息、学术、产业、金融甚至政府机构的广泛关注,推动了大数据技术和应用的空前繁荣。在交通运输物流领域,我国不仅已经成为一个交通大国,同时也是交通数据大国。

为有效实现对交通运输物流大数据的整合与应用,以及对数据产品及软件产品的增值,针对交通运输物流数据特点及对大规模数据离线处理和实时分析的迫切需要,本书对大数据分析平台软件架构方案进行了设计,对大数据环境下交通运输物流领域数据资源存储、计算和资源调度管理、离线数据并行计算和在线数据实时计算等方法进行剖析,实现了实时热力分析、车货信息资源实时匹配、热点词分析、物流企业信用评价等典型的交通运输物流领域大数据应用,受限于数据的规模和来源的局限性,本书提出的相关方法和模型在实际应用中还需要根据实际应用场景进行适当的改进,相关架构设计方案、数据分析方法和模型具有一定的理论指导意义和参考价值。

本书的主要研究内容如下。

(1) 为了满足交通运输物流领域大数据的离线与实时分析的需求,设计了一种大数据平台软件基础架构,该架构主要由分别负责离线数据分布式存储和并行计算的 Batch Layer、为不同的应用提供高速用户数据视图的 View Layer 以及负责为流式数据提供消息队列并进行实时计算的 Speed Layer 组成,该软件基础框架可以应用在不同的大数据环境下,提高大数据的分析能力与开发效率。

(2) 结合分布式实时计算框架提出了一种基于 CluStream 流式聚类算法的交通运输物流活动区域热度实时分析方法,该方法针对流式数据实时性强和无限流入等特点,对运输车辆活动数据流基于“金字塔”时间窗口方法进行抽样与聚类分析,得到不同区域的物流活动的热度指标,该方法不仅能够解决无限流入数据存储空间效率低下的问题,也能够满足交通运输物流流式数据在进行相关分析的实时性要求。

(3) 为了提高物流公共信息平台中货运供需信息资源的匹配效率,本书结合分布式实时计算框架,建立了一种车货供需匹配数学模型,并使用改进的量子进化算法对此问题求解,实验结果表明量子进化算法表现出更好的收敛速度、准确性和稳定性,可以高效地搜索到较为优秀的车货信息资源匹配方案,为车主和货主推荐较

为合理的车货供需信息资源。

(4) 结合分布式并行计算框架和非结构化文本分析技术,提出了基于 MapReduce 的词频统计与文本分类方法,并应用在交通运输物流热点词分析和交通运输物流新闻资讯分类导览系统中,该方法能够有效提高管理决策者对交通运输物流领域信息和知识的收集分析能力。

(5) 为了在大数据背景下更全面地分析物流企业的信用,引入了第三方评价好评率和涉案胜诉率等新的信用评价指标并结合传统指标建立了物流企业信用评价体系,通过网络爬虫收集各级人民法院公开公示的法律文书文本内容,从中抽取案件相关信息用于统计相关物流企业的涉案胜诉率,并提出了基于 MG 算法的第三方评价特征抽取方法,解决了大规模文本特征分析效率低下的问题。该方法对于构建科学、有效和可行的物流企业信用评价体系具有一定的现实意义。

随着交通运输领域的信息化发展和大数据技术的日新月异,相关技术方法仍然有很大的发展空间,本书的内容在今后的研究中需要进行不断改进,敬请广大读者多提宝贵意见。

本书受教育部人文社会科学研究项目(18YJC630124)和辽宁省教育厅科技研究项目(L2014203)资助出版。本书内容第1章由大连科技学院蒋晶晶编写,其余内容由大连海事大学牟向伟编写。

目 录

1 绪论	(1)
1.1 大数据产生的背景	(1)
1.2 大数据的发展历程	(2)
1.2.1 国外大数据的发展历程	(2)
1.2.2 国内大数据的发展历程	(3)
1.3 我国交通运输物流发展情况	(4)
1.3.1 我国交通运输物流系统建设情况	(4)
1.3.2 我国交通运输物流信息化发展情况	(5)
1.4 国内外交通运输物流大数据研究进展	(6)
1.4.1 国外交通大数据研究进展	(6)
1.4.2 国内交通大数据研究进展	(7)
1.4.3 大数据在交通运输物流行业中的应用	(8)
1.5 交通运输物流领域数据分析的问题	(9)
1.6 本章小结	(11)
2 大数据分析平台软件技术基础架构	(12)
2.1 大数据基础技术框架简介	(12)
2.1.1 分布式存储	(12)
2.1.2 分布式计算	(13)
2.1.3 实时计算	(14)
2.1.4 资源调度管理	(15)
2.2 大数据基础软件技术架构设计	(16)
2.2.1 设计原则	(17)
2.2.2 大数据分析平台软件基础架构设计方案	(19)
2.3 本章小结	(49)
3 交通运输物流大数据的数据组织与描述	(51)
3.1 泛数据生态圈	(51)
3.2 大数据资源的存储模式	(53)
3.2.1 大数据特征	(53)

3.2.2	交通运输物流大数据的数据存储模式	(55)
3.3	基于本体的交通运输物流领域知识描述方法	(58)
3.3.1	相关理论	(59)
3.3.2	交通运输物流领域大数据本体的构建	(63)
3.4	本章小结	(68)
4	交通运输物流领域流式数据实时分析与应用	(69)
4.1	流式数据实时分析方法	(69)
4.1.1	流式数据的定义	(69)
4.1.2	流式数据的特点	(70)
4.1.3	流式数据抽样统计方法	(70)
4.1.4	流式数据聚类方法	(73)
4.2	流式数据实时分析在交通运输物流领域中的应用	(77)
4.2.1	应用1:基于CluStream的交通运输物流活动热点实时分析	(77)
4.2.2	应用2:大规模流式数据环境下车货实时匹配分析	(83)
4.3	本章小结	(98)
5	交通运输物流领域离线数据并行计算与应用	(99)
5.1	基于MapReduce的交通运输物流热点词频统计方法	(99)
5.2	基于非结构化文本分类的交通运输物流资讯新闻导览方法	(105)
5.2.1	文本分类技术综述	(106)
5.2.2	文本训练	(112)
5.2.3	文本分类	(113)
5.2.4	交通运输物流资讯分类导览系统的实现	(114)
5.3	大数据环境下的物流企业信用评价方法	(118)
5.3.1	应用背景与意义	(118)
5.3.2	物流企业信用评价方法	(119)
5.3.3	大数据环境下的物流企业信用评价指标	(124)
5.3.4	数据收集系统	(132)
5.4	本章小结	(139)
6	总结	(141)
	参考文献	(144)
	附录	(154)
附录A	“法律文书”文本分析源代码	(154)
附录B	物流企业信用评价数据表(部分)	(161)
附录C	信用等级与对应评分表	(166)

1 绪 论

1.1 大数据产生的背景

随着新技术和新理念的应用,数字化已经渗透到我们生活的方方面面。每个人都是大数据的“生产者”,每天都在产生和共享大量的数据。根据美国著名市场调研公司 IDC 的统计,全球每秒钟发送的电子邮件达 290 万封。如果在一分钟内读完一本书,就足够一个人日夜阅读 5.5 年。每天有 28800 小时的视频上传到 YouTube 上,足够一个人日夜观看 3.3 年;每天有 5000 万条信息发布在 Twitter 上。如果你在 10 秒内读完一条信息,足够一个人夜以继日地浏览 16 年。亚马逊每天产生 630 万订单,谷歌每天处理 24PB 的数据,而大数据仍然每 18 个月使全球信息量翻一番。

在中国,数据的规模也是巨大的。截至 2013 年 12 月,新浪微博月活跃用户 1.291 亿,日均活跃用户 6.14 亿,每天发布和转发的信息超过 2 亿条。截至 2013 年底,淘宝拥有近 5 亿注册用户,每天有超过 6000 万的固定访客。与此同时,每天的网上商品数量已超过 8 亿件,平均每分钟售出 4.8 万件商品。2015 年 11 月 11 日电商购物节,支付宝单日交易 7.1 亿笔,是 2014 年同日的 6.7 倍,支付宝交易数为 1.058 亿笔,8.59 万笔/秒的峰值为 2014 年的 2.23 倍(2014 年 11 月 11 日的峰值为 3.85 万笔/秒)。

随着移动通信的发展和智能手机的普及,移动互联网得到蓬勃发展,移动互联网的发展使更多的人成为数据的生产者。根据相关统计报告,2016 年全球互联网用户数将达到 32 亿人,约占全球总人口数的 44%;其中,移动互联网用户总数将达到 20 亿。相关报告预计,如果没有新型联网设备出现,截至 2020 年移动互联网用户总数将以每年 2% 的比例增长。中国互联网信息中心(CNNIC)发布的第 37 次《中国互联网络发展状况统计报告》指出,到 2015 年 12 月,中国的网民数量已经达到 6.88 亿人,其中移动互联网用户的数量已经达到 6.2 亿人,占 90.1%,与 2014 年底相比增长了 4.3 个百分点,其中仅通过手机上网的网民占 18.5%,较 2014 年底提升了 3.2 个百分点。在国外,据估计用户每月在 Facebook 上花费 7000 亿分钟,移动互联网用户发送和接收的数据量为 1.3EB。在我国,中国联通用户上网记

录条数为 83 万条/秒,即 1 万亿条/月,对应数据量为 300TB/月,或 3.6PB/年。

大数据被一些人定义为“21 世纪的新石油”。2010 年个人用户产生的数据达到 1.2 ZB,相当于全世界海滩上的沙子数量的总和,这意味着 TB、PB、EB 已经过时,全球正式进入数据存储的 ZB 时代(即“泽它时代”),进入大数据时代。根据 IDC 预测,到 2020 年,整个世界的的数据总量将比 2010 年增长 30 倍,达到 35.2ZB。而根据 IDC 在 2012 年 12 月份对中国数字世界的预测,到 2020 年,中国数据总量将增长 24 倍,达到 8.6ZB。

庞大的数据量给我们带来了前所未有的挑战,例如,如何收集大数据,去伪存真,保证数据的全面性和可靠性?如何有效地存储和处理海量的数据?如何从大量数据中挖掘出隐藏的巨大商业价值?如何对结果进行可视化呈现?如何使现有系统弹性地应对数据和负载的动态变化?如何快速构建大数据处理系统并且保证系统的安全简便可用?目前,这些问题已经引起了信息、学术、产业、金融等各个行业甚至政府机构的广泛关注,人们逐渐开始研究如何运用大数据理念和技术来解决上述问题,从而推动了大数据技术和应用的空前繁荣。

1.2 大数据的发展历程

1.2.1 国外大数据的发展历程

(1) 2005 年 Hadoop 项目诞生。从技术上看,Hadoop 主要包括两部分,一是对海量数据进行存储的分布式文件系统 HDFS,二是对海量数据进行分析的 MapReduce 并行计算框架。Hadoop 从技术层面上搭建了一个使对结构化和复杂数据快速、可靠分析变为现实的平台。

(2) 2008 年 12 月,美国“计算社区联盟”发表白皮书《大数据计算:在商务、科学和社会领域创建革命性突破》,详尽阐述了大数据对社会治理的推动作用以及潜在的商业价值。美国“计算社区联盟”可以说是最高提出大数据概念的机构。

(3) 2011 年 5 月,著名咨询公司麦肯锡发布了《大数据:创新、竞争和生产力的下一个前沿领域》报告,这是专业机构第一次全方面地介绍和展望大数据,在这份报告中“大数据”概念得到清晰阐述。报告提出“大数据时代已经到来”,并详细列举了大数据的核心技术,深入分析了大数据在不同行业和业务职能领域的应用,认为只要给予适当的政策支持,“大数据”将促进生产力增长并推动创新。

(4) 2011 年 6 月,全球互联网巨头已意识到“大数据”时代数据的重要意义,包括 EMC、惠普、IBM、微软等 IT 巨头纷纷通过收购数据相关厂商来实现技术整合,

部署“大数据”战略。

(5) 2012年1月,在瑞士沃达斯论坛上发布了《大数据,大影响》的报告。该报告认为大数据是一种新型的经济资产,并从金融服务、健康、教育、农业、医疗等多个领域阐述了大数据给世界经济社会发展带来的机会。

(6) 2012年3月,奥巴马政府投资2亿美元启动“大数据研究和发展计划”,以推动大数据的收集、储存、管理、分析和相关技术问题的研究。

(7) 2012年7月,联合国发布了《大数据促发展:挑战和机遇》,这是一份关于大数据政务的白皮书,总结了各国政府如何利用大数据更好地服务和保护人民。

(8) 2013年,英国政府发布《英国数据能力发展战略规划》,旨在利用数据产生商业价值、提振经济增长。

(9) 2014年5月,美国白宫发布了2014年全球“大数据”研究报告《大数据:抓住机遇、保存价值》。报告阐述了大数据发展所带来的机遇与挑战,并从技术的视角对大数据技术与隐私保护之间的关系进行了深入分析,探讨了可以采用哪些技术加强隐私保护。

(10) 2014年,欧盟在发布了《数据驱动经济战略》,旨在利用大数据改造传统治理模式,大幅降低公共部门成本,并促进经济增长和就业增长。

(11) 2015年,国际科学理事会发布《Open Data in a Big Data World》协议。该协议认为“大数据”已经成为科学发现的重要机会,而“开放数据”将提高公共研究企业的效率,生产力和创造力,并抵消知识私有化的趋势。

(12) 2016年5月,白宫发布了《联邦大数据研发战略计划》报告,在已有的基础上提出美国下一步的大数据发展战略。通过一系列的改革措施,美国大数据战略形成了跨部门协同工作的机制。

(13) 2017年9月,医疗保健研究与质量局发布美国首个可公开使用的数据库,其中包括全美600多个卫生系统。白宫科技政策办公室一直积极与他国展开合作,以预防数字经济监管障碍、促进信息流动和反对数字本地化等。

(14) 2018年4月,英国专门发布《工业战略:人工智能》报告,立足引领全球人工智能和大数据发展,从鼓励创新、培养和集聚人才、升级基础设施、优化营商环境以及促进区域均衡发展等五大维度提出一系列实实在在的举措。

1.2.2 国内大数据的发展历程

(1) 2008年11月,由60余位大数据技术爱好者在中科院首先举办了“Hadoop in China”第一次会议。在国家863“中国国家网络软件研究与开发”课题的支持下,Facebook、Yahoo!、Baidu、MySpace、Alibaba、FranceStremezzo、Nortel等企

业的科研人员和相关人员进行了一场关于“Hadoop 在实际使用中的经验以及心得”交流会,初次掀起了大数据技术应用的神秘面纱,是我国对大数据技术进行的首次探索。

(2) 2012年7月,国务院印发了《“十二五”国家战略性新兴产业发展规划》。该规划明确提出支持大数据存储、处理技术的研发与产业化。

(3) 2012年7月,在北京大学举行了“首届中国大数据应用论坛”,会议的主要议题包括大数据的发展趋势、不同场景的大数据应用、云计算与大数据、大数据与商业智能等。

(4) 2012年10月,中国计算机学会大数据专家委员会成立。

(5) 2015年7月,国务院颁布《关于运用大数据加强对市场主体服务和监管的若干意见》,为充分运用大数据先进理念、技术和资源,加强对市场主体的服务和监管,推进简政放权和政府职能转变,提高政府治理能力给出了指导意见。

(6) 2015年9月,国务院发布《促进大数据发展行动纲要》,旨在促进中国数据技术的发展。纲要对我国大数据发展进行了顶层设计和统筹规划,彰显和强化了大数据在国家战略中的突出地位。

(7) 2016年3月,在国家“十三五”规划中提出实施大数据战略,促进数据资源开放共享。大数据产业第一次明确出现在规划中。

(8) 2017年1月,工业和信息化部印发大数据产业“十三五”发展规划。

(9) 2018年1月,国家发改委宣布了政务信息系统整合共享工作最新进展,已有71个部门、31个地方实现了与国家共享交换平台的对接。

(10) 2019年9月大数据产业生态联盟联合赛迪顾问发布《2019中国大数据产业发展白皮书》,报告指出2018年中国大数据产业规模为4384.5亿元,预计2021年将达8070.6亿元。

1.3 我国交通运输物流发展情况

1.3.1 我国交通运输物流系统建设情况

改革开放40多年来,我国交通物流行业实现了跨越式发展,已建成世界最大的交通运输产业体系。

在公路运输方面,到2020年,全国公路里程已达到520万千米,基本建成全国公路网。高速公路总里程已达到16万千米(世界高速公路总里程约50万千米),位居世界第一,比美国高出5.3万千米。铁路网全长14.63万千米,仅次于美国的25

万千米,是世界第二大铁路网。3.8 万千米的高铁网络是世界首个高铁网络。拥有世界十座最长跨海大桥中的 6 座,目前在桥梁和公路隧道技术方面处于世界领先地位。中国的交通运输系统承载着世界上最大的客运量和货运量。公路客运量平均每年 190 亿人次,相当于每人 12 或 13 人次。2019 年,铁路旅客发送量达到 36.6 亿人次,货物发送量超过 300 亿吨。

航空运输方面,中国拥有 241 个机场,客货运量居世界第二位。随着人民生活水平的不断提高,我国民航客货运量迅速增长,越来越多的人开始使用航空出行。

水运方面,世界前 10 大港口中,中国占了 8 个,占吞吐量的 80% 以上,其中宁波舟山港排名第一。在世界十大集装箱港口中,中国占 7 个,占吞吐量的 70%,上海港居世界第一。长江、西江、黑龙江等可通航核心航道在内的内陆走廊长度达 12.77 万千米,居世界首位。

在城市交通方面,中国有 7 万多条公共汽车和有轨电车线路,总长约 140 万千米。轨道交通虽然刚刚起步,但发展迅速,中国现有轨道交通线路 226 条,总里程 7000 多千米。目前,轨道交通正从一些核心国际大都市向一、二线城市延伸。每年城市交通出行超过 1000 亿人次,居世界首位。经过几十年的发展,中国已经成为一个交通大国。

1.3.2 我国交通运输物流信息化发展情况

交通物流信息化与大数据密切相关。我国在沿海港口、远洋航运、高速铁路等领域的信息化水平已进入世界先进行列,而在空中交通管理体系和沿海内河通信等领域,与世界先进水平的差距也在不断缩小。国内许多大数据公司在高速公路网络收费、通信、监控等方面进行了大数据挖掘,为发现高速公路交通数据背后隐藏的、未知的规律性知识提供了依据。

此外,铁路和民航旅客票务信息服务已达到世界先进水平。12306 火车票网上订票系统、携程、去哪儿网等机票订票系统为人们出行带来了极大的便利。北京、上海、广州、深圳等城市已经建立了较为完善的城市智能交通系统。上海世博会通过分析每天进出上海的所有交通流量和客流,并预测第二天世博园内的人数,为世博园提供帮助。在综合交通信息化领域,通过海铁多式联运工程、综合客运枢纽工程等信息化示范试点项目,我国综合交通信息化取得阶段性突破。如南京南站公路铁路一体化枢纽,通过多种交通方式的协调联动,实现了一体化运输,并取得了良好的效果。

目前,我们已经进入“互联网+”时代,“互联网+”物流、“互联网+”交通、大数据已成为国家战略,将对未来整个交通物流行业的信息化、智能化发展产生深远影响。

1.4 国内外交通运输物流大数据研究进展

1.4.1 国外交通大数据研究进展

美国 INRIX 公司与新泽西运输部展开交通大数据方面的合作。INRIX 通过汽车和手机上的 GPS 设备上的信号收集主干道上的车辆速度数据,向新泽西州交通运输部发出主干道上的危险情况实时警告。同时向司机的 GPS 设备或手机发送警告,提醒司机注意道路危险情况。

2010 年英国伦敦为了将伦敦市的交通信息公开,成立了 TransportAPI 公司,该公司将交通数据公开,并使包括机场、轨道服务商以及市政府都可以采集、审查、整理这些交通数据,为伦敦市的交通大数据奠定了基础。

西班牙电信于 2012 年 10 月成立了名为“西班牙电信动态洞察”的大数据业务部门,推出名为“智慧足迹”的商业服务。甚至还计划面向不同行业推出系列服务产品,如具有交通流量管理功能的“智慧城市”。

2014 年 5 月 29 日,日本丰田汽车宣布通过收集汽车的位置和速度等庞大的“大数据”,开发出可提供交通量和行驶线路等交通信息的服务。丰田的“大数据交通信息服务”使用通信信息终端,收集到 330 万台搭载导航系统的汽车行驶 1 年(绕地球 83 万圈)的位置和速度等信息,通过分析这些信息为用户提供在地图上显示道路的拥堵情况、发生灾难时可行驶的道路以及避难所的位置等服务。

目前韩国企业对大数据的研究已经不仅仅停留在本产业领域内,融合交通数据、流动人口数据等跨领域的大数据分析,也成为商家们打造精准营销的手段之一,据了解,韩国利用大数据所创造的市场交易规模平均每年以 35% 的速度增长,韩国政府已于去年成立了韩国大数据分析中心,对韩国交通领域、经济领域等 45 种领域的公共基础数据进行分类整理,为市民的日常生活和企业的经济活动提供了参考依据。

2015 年 4 月 27 日新加坡陆路交通管理局局长 Chew Men Loeng 在第十四届亚太智能交通论坛上提出了大数据在新加坡交通领域中的应用,探讨分析了在大数据时代,根据新加坡的情况,怎么样获取大数据,怎么样为新加坡的未来交通的生态系统进行设计。

拥有世界第二多人口数的印度,近两年也持续运用大数据,期许打造出一个更便利、更近民的智慧城市。印度的一些新城市将会运用大数据,建立一套地理资讯系统。

欧盟将构建交通大网络,搭建物流生命线。欧盟交通网络存在在快速建设高速公路的同时,道路交通却仍旧日益拥挤,路网通行能力不能满足交通量增长的需要等问题。为此,从 20 世纪 70 年代末开始,在欧盟主导下,各国开始了对“公路智能运输系统”(ITS)的研究与规划,至 90 年代中期开始进入调试与实用阶段。ITS 系统以强化交通安全、提高运输效率、节约能源消耗、保护自然环境和增加舒适性等为关键环节。

“阿里云”将建迪拜数据中心,新的合资公司总部设于迪拜,双方将共同建设一座数据中心,为迪拜的交通、通信、城市基础设施、电力、经济服务和城市规划等六大支柱领域提供服务。

2016 年 6 月新西兰交通管理局按照历史的交通大数据,进行挖掘分析,总结出了一些城市路段的拥堵点为市民出行提供参考。

1.4.2 国内交通大数据研究进展

伴随着交通管理信息系统的不断建设、完善与发展,交通运输物流领域积累了大量宝贵的数据资源。为了适应大数据时代的发展要求,更好地利用大数据为交通现代化服务,国内学者对大数据方法和技术在交通领域(如,智慧交通、道路交通运输和物流运输体系等方面)的应用进行了广泛而深入的研究。

陈美提出大数据在公共交通中的应用,结合英国对大数据在交通管理中的应用案例,分析大数据管理可以解决及时、高效、准确的交通数据获取问题。为大数据在交通领域的应用做了个初步分析。

2013 年北京市交通委员会顾涛结合大数据提出了治理北京城市交通拥堵的策略。他认为大数据、物联网时代的来临,为科学施策治理城市交通拥堵提供了条件,通过对科学收集的海量数据进行针对性分析和归纳,并提炼出有价值的信息,可以为科学决策提供依据和支撑。大数据技术可以实现城市交通的智能化。

刘伟杰和保丽霞探讨了交通大数据对构建综合交通运输物流体系的支撑作用及实施关键,提出了大数据在整个交通运输物流体系中的作用主要体现在对管理者、出行者、社会环境 3 类对象 7 个方面。并且结合了上海综合交通运输物流体系的现状,从规划、机制以及全寿命 3 个方面探讨了交通大数据在上海市构建世界一流运输体系中应用。

段宗涛和康军等人将大数据应用到车联网的研究中,提出了一种车联网大数据环境下的交通信息服务协同体系。他们的新型交通信息服务协同体系将交通信息服务模型描述、基于机会通信协议、语义交通信息匹配算法和上下文驱动四个部分组合,建立了车联网大数据环境下的新型交通信息服务协同原型系统。

熊刚、董西松等人重点分析总结了城市交通大数据的若干研究内容及核心技术,提出了城市交通大数据智能应用系统解决方案和几种典型应用,在城市交通和智慧城巯领域的大数据研发和应用领域进行了初步探讨。

王雅琼、杨云鹏、樊重俊思考利用大数据技术改进智慧交通服务、缓解交通拥堵问题,分析智慧交通对大数据应用的需求变化,探讨智慧交通中大数据应用的新模式,从而利用大数据更好地为人们提供智能化交通服务。

段宗涛、郑西彬等人对道路交通领域业大数据进行了相关研究。他们分析归纳了道路交通大数据的来源并分类;而后分析了道路交通大数据处理和管理的支撑技术;获得了道路交通大数据的关键问题以及开展道路交通大数据研究所需的技术基础和道路交通大数据在未来可能的应用,阐明了道路交通大数据研究的关键技术及其相关应用。

闫俊伟、凌卫青和王坚结合本体技术对交通大数据进行知识语义描述,结合大数据技术提出了一种基于本体的交通大数据分析框架。该框架中对交通以路网拓扑、道路交通对象和道路交通信息三个层次进行语义关系描述,制定映射文件,构建城市道路交通本体库,为大数据分析的多源多维数据关联分析及知识挖掘提供语义查询支持。该框架能够根据交通分析需求快速有效地找到目标数据,在大数据分析 with 大数据存储之间起到了逻辑关联的作用,对现有交通数据分析具有重要意义。

1.4.3 大数据在交通运输物流行业中的应用

近年,大数据的相关技术方法在交通运输物流领域的应用越来越多,在车货匹配、运输线路分析、销售预测与库存、设备修理预测、物流中心选址等方面有着广泛的应用。

1. 大数据技术可以改变运输资源匹配模式

物流运输行业具有车多货少的特点,供过于求,导致了运输卡车等待货物和交付货物的大部分时间被浪费,无形中增加了企业的经营成本。随着移动互联网、云计算、大数据等新技术的发展,物流信息平台迎来了新的机遇。目前,市场上出现了大量的车货匹配信息平台 and 移动端应用。例如,仅 2014 年,市场上就有 200 多个车货匹配应用。通过对车源信息和货源信息进行大数据分析,可以使车主的货物与货主的运力需求之间产生高效的匹配,从而降低货物运输车辆的返程空载率,减少资源浪费。此外,大数据技术还可以解决信息平台上无货或虚假供应的问题。

2. 基于大数据技术的运输路线优化

UPS 司机每天要递送 120 到 175 件货物。如果每位司机每天少开一英里,公

司可以节省 5000 万美元。为了找到任何两个目的地之间最有效的路线,UPS 使用大数据技术开发了“猎户座”系统(on-road integrated optimization and navigation),因为司机可以在多条路线之间进行选择,该系统可以实时分析 20 万条可能的路线,并在大约 3 秒内找到最佳路线。到 2013 年底,“猎户座”已经在大约 10000 条线路上使用,为公司节省了 150 万吨燃料和 14000 立方吨二氧化碳排放。

此外,UPS 通过大数据分析,规定车辆在右路行驶到任何送货地点时都应避免左转。原因在于 UPS 通过大数据分析技术实现的左转增加了车辆事故风险,影响了其效率。到 2012 年,由于右转规则和其他改进,UPS 已经节省了大约 1000 万加仑的燃料,减少了相当于 5300 辆汽车一年的排放量。

3. 利用大数据技术预测销量和库存

使用大数据技术通过分析类别可以确定哪些可以用来促进商品、商品用于排水,同时,根据以往销售数据建模确定当前安全库存商品,当库存接近安全库存时进行早期预警,及时不断优化库存结构,降低库存成本。

4. 大数据技术可用于预测运输工具维护

车辆维修人员需要不定期对车辆进行检查,只在需要时根据数据分析进行检查,通过监控车辆的每个部件,UPS 现在只需更换需要更换的部件,从而节省了维护成本。如使用无线远程信息处理系统在运载工具负载传感装置来检测特定部分的性能,如车辆运行状态、速度、方向、刹车和引擎,并收集多种指标反映卡车,的速度、行驶时间和油压,以有效降低成本,提高驾驶安全性。

5. 物流中心选址应用大数据技术

物流中心选址的选址直接影响着物流中心活动的成本,所以物流中心选址必须充分考虑物流中心本身的特点、商品特点和交通条件等因素。而大数据分析技术可以很好地解决这个问题。

1.5 交通运输物流领域数据分析的问题

尽管大数据的应用意味着大机遇,拥有着巨大的商业价值,但将大数据技术应用到交通领域还存在以下一些问题。

1. 大数据的质量和时效性难以把握

交通大数据可能来源于不同的交通机构,数据结构和数据完整性也可能随着数据源的不同而不尽相同,而且在数据收集阶段,由于数据变化较快,有效期很短,这可能导致所收集的数据无效,从而在一定程度上影响数据的质量。因此,对于交

通领域相关部门或企业而言,从多个数据源及时获取高质量的数据并进行有效整合,是一个巨大挑战。

2. 部分行业对大数据技术缺乏高度的重视和支持

大数据在中国还处于不成熟的阶段,再加上大数据本身的多样性和复杂性,使得大数据的质量就无法得到有效、全面的保证,部分行业(如许多的物流企业高层管理人员)还没有意识到大数据挖掘技术、大数据分析技术给自身企业带来的商业价值到底有多大,对大数据的认识还没有真正提升到企业发展的战略高度。因此,应加强对大数据的认识,清楚大数据在信息时代的真正价值,建设完善的数据中心和完善的数据质量保证制度,促进行业或企业的发展。

3. 物流企业高层管理者对大数据技术缺乏高度的重视和支持

只有得到了物流企业高层管理者的重视,一系列跟大数据有关的应用及发展规划才能有望得到推动,大数据的价值才能在物流的运营过程中真正地挖掘出来。然而,大数据在中国还处于不成熟的阶段,再加上大数据本身的多样性和复杂性,使得大数据的质量就无法得到有效、全面的保证,许多的物流企业高层管理人员还没有意识到大数据挖掘技术、大数据分析技术给自身企业带来的商业价值到底有多大,对大数据的认识还没有真正提升到企业发展的战略高度。因此,物流企业高层管理者应当加强对大数据的认识,清楚大数据在信息时代的真正价值所在,建设完善的数据中心和完善的数据质量保证制度,带领企业迎接这场没有硝烟的大数据战争。

4. 数据中心亟需专业的数据管理人员

专业数据管理人员的配备才是保证大数据质量的关键,由于大数据本身的多样性、复杂性增加了大数据在处理和和管理上的难度,因此,在大数据环境下,交通部门亟需专业的数据管理人员。

5. 将非结构化的数据转化为结构化的数据是一项巨大挑战

数据有着结构化数据和非结构化数据之分,结构化数据是指储存在数据库里,能用二维表结构来表达的数据;而非结构化数据是指包括所有格式的文本、图片、办公文档、各类报表 html、xml、图像和音频/视频信息等等。对于交通机构来说,将非结构化数据转化为结构化数据是一项巨大挑战。

6. 数据开放程度、共享程度低,共享机制不完善

交通部门拥有大量的数据资源,但利用率不高,这不仅造成了资源的浪费,也给人们的生活带来了很大不便。例如,我国城市公交一卡通已覆盖了约 250 多个地级以上城市,但因各地公共交通一卡通标准不一、各城市间不能联通、无法实行