

交通运输大数据应用进展

杨琪¹, 刘冬梅²

1. 交通运输部公路科学研究院, 北京 100088
2. 国家智能交通工程技术研究中心, 北京 100088

摘要 随着交通运输服务与管理信息化、智能化的发展, 行业内集中了大量丰富的数据资源。分析了交通运输大数据的特点、类型及数据量, 简述了综合交通出行大数据开放云平台的构架、功能和应用效果。通过研究交通运输大数据在交通拥堵治理、高速公路智能稽查、客流统计及民航空管等方面的应用案例, 指出交通运输大数据发展及应用趋势混杂性更强, 分析处理结果可能会有相互矛盾的情况, 随着数据量的增加, 不精确性对整个分析结论的影响力变小; 当数据量扩大后, 简单算法可能更加实用等。

关键词 交通运输; 大数据; 开放云平台

随着计算机数据存储、处理技术的发展, 以及有线通信和无线通信系统的大量应用, 尤其是数据传输、处理、存储成本的降低, 大数据的采集及利用得到各行业的重视^[1]。中国大中型城市的交通问题日益突出, 例如, 很多一线及二线城市交通拥堵严重, 事故频发, 造成大气污染的问题也不容忽视, 需要对交通运输获取实时信息, 进行科学管理, 提高交通运输效率, 在此背景下, 大数据在交通领域的应用成为热点^[2]。

交通运输数据存在容量大、类型多、数据来源五花八门、结构化数据和非结构化数据分布不均、数据关联性强等特点, 使交通大数据在发展过程中, 演化出数据处理技术手段多样、对大流量实时数据处理要求高、大数据分析面向多用户、多用途等特征。随着交通运输服务与管理信息化、智能化的发展, 在行业内集中了大量丰富的数据资源, 基本具备了大数据的特征^[3], 交通运输行业的大数据应用在所有行业领域应用的关注度中名列前茅。互联网企业是交通数据的重要来源, 以

摩拜单车和滴滴出行平台为例, 摩拜仅450万辆智能共享单车每天就可以产生超过5 TB的骑行数据, 而滴滴每天处理的数据量超过2 PB^[4]。

交通运输的数据主要可以分为两大类。第一类为交通运输运行所直接产生的数据, 包括道路、桥梁、铁路、航线、机场、巷道、港口、码头等基础设施数据, 各类交通流数据, 运载工具本身及运行数据, 交通控制及管理数据, 各类视频、图片等数据。据不完全统计, 目前交通运输行业业务信息管理系统约有700余种, 数据规模以PB计。例如, 全国重点营运车辆联网联控系统每天接受的300多万辆客货运车辆的10亿多条卫星定位数据; 海事系统的船舶自动识别系统(automatic identification system, AIS)月均数据量达到17.2亿条; 高速公路运营企业采集传输的视频数据日均数据量已达到PB级, 联网收费数据月均7亿条以上。第二类为交通运输相关行业和领域导入的数据, 例如气象、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

收稿日期: 2017-09-18; 修回日期: 2018-12-25

作者简介: 杨琪, 研究员, 研究方向为智能运输系统, 电子信箱: yangqi@itsc.cn

引用格式: 杨琪, 刘冬梅. 交通运输大数据应用进展[J]. 科技导报, 2019, 37(6): 66-72; doi: 10.3981/j.issn.1000-7857.2019.06.009

的交通状况数据,例如通过微博、微信、论坛、广播电台等提供的文字、图片、音视频等数据。

目前,第一类数据的标准化程度相对较高,管理的协调程度也较高。而对于第二类数据,由于数据分散、形态各异,用传统技术难以在合理的时间内进行管理、处理和分析。

1 综合交通出行大数据开放云平台 ——“出行云”

交通运输数据的拥有者主要有2类,即各级交通运输行业主管部门和企业。

在“互联网+”便捷交通中最重要的内容之一就是:交通运输主管部门和企业将服务性数据资源向社会开放,通过互联网平台为社会公众提供实时交通运行状态服务,包括查询、出行线路规划、网上购票、智能停车等服务^[5-6]。

在中国交通运输部的支持下,交通运输部公路科学研究所和北京百度网讯科技有限公司共同打造了“基于开放式综合交通出行服务信息应用云平台”,通过创新政企合作模式、促进行业与互联网出行服务信息的开放共享,为社会公众提供更丰富、更权威、更高品质的出行信息服务。平台总体架构如图1所示。

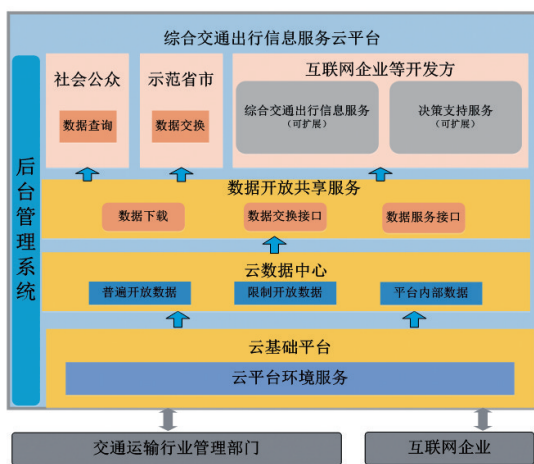


图1 综合交通出行开放云平台构架
Fig. 1 Architecture of open cloud platform of integrated transportation

平台的主要功能和特点:云平台的交换功能向所有社会企业开放,在规范的共享、交换机制下,将全国范围分散的综合交通数据进行整合;充分利用云平台

的用户资源以及互联网企业多领域信息服务产品线,通过增加服务出口、名称互相标识、多媒体宣传等方式进行协同;通过深度发掘云平台的综合交通运输数据,为交通行业管理部门提供数据统计、分析、预测等行业管理所需的相关决策支持。出行云平台开放的数据类型如图2所示。

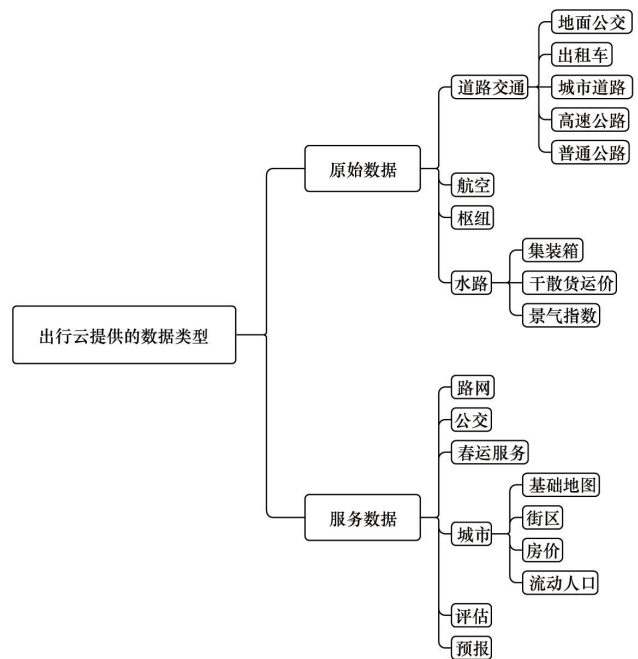


图2 出行云平台开放数据的类型
Fig. 2 Types of open data on the travel cloud platform

目前,江苏、辽宁、重庆、四川、河南、吉林、京津冀、黑龙江、湖北等十几个省市先后加入,开展了以平台为基础的全国综合交通出行服务信息共享政企合作模式的应用示范。目前,接入平台的数据类型包括基础设施数据、几十万辆客运车辆的数据、出租车实时定位数据、公交车辆数据、公路网络运行数据、高速公路阻断数据、路况视频截图等。

云平台体系的建立,打破了各省市间、各行业间、政府企业间的信息孤岛,完成了跨区域、综合交通数据的整合与共享^[7];基于云平台所提供的功能接口,江苏省、辽宁省、重庆市等改造升级了当地出行信息服务系统,为百姓提供实时交通运行状态查询、跨区域跨交通工具的出行路线规划等功能,推进了基于互联网平台的多种出行方式信息服务对接和一站式服务;信息服务由具有技术和受众优势的企业提供,降低了交通行业管理部门的相关建设和运营费用。

2017年夏天,中国长江中下游沿江地区及江淮、西南东部等地出现入汛以来最强降雨过程,全国有26省(市、区)1192县遭受洪涝灾害,暴雨给人民群众生活造成了极大影响和不便。江苏省通过本平台与百度公司合作,将因暴雨阻断的道路及时醒目地标示到百度地图上,并实时调整客户端的导航线路,帮助用户提前采取绕行方案,受到出行者的欢迎。

平台结合专题分析,利用定位数据和智能化分析技术,连年开展全国春运“黄金周”期间的交通分析工

作。2016年春运期间,全国春运服务调查问卷累计点击量达511万,回收有效问卷62.3万份;采集到百度日均260亿次定位请求数据,获取了11.63万条城市对之间出行信息;舆情监测涵盖了新闻报道、微博评议、论坛网帖、微信公众号等共计87.3万条信息。基于大数据的调查结果真实反映了旅客出行目的、出行方式、出行时间与空间分布等特征。在此基础上,进行跨省(市、区)出行及省内出行流量流向分析(图3),为今后春运的运力安排提供了科学的基础。

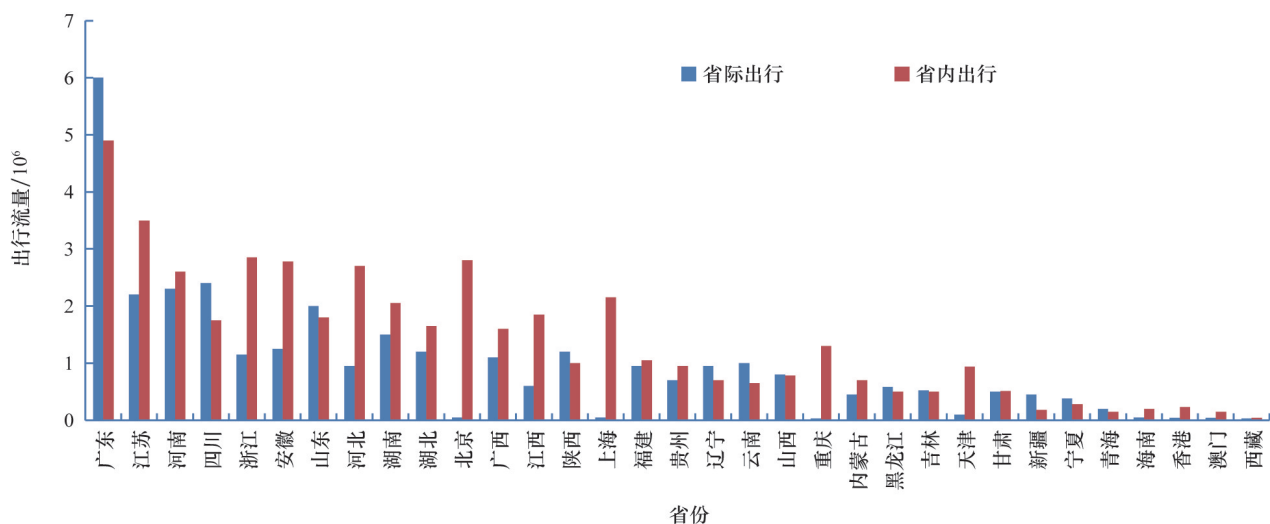


图3 2016年跨省出行及省内出行流量分析

Fig. 3 2016 travel flow analysis of inter-provincial and travel within the province

目前,云平台有23个省市(区)交通运输厅/局、32家企业、12家高校科研机构、1家媒体等共68家单位作为联席成员单位,向交通运输行业管理部门提供了近100项支持海量用户调用的出行服务开发基础接口,以及实时路况、实时公交、地图路网、室内地图4项共建服务接口,并引导互联网公司与交通行业管理部门建立数据资源置换技术资源的合作模式。

2 交通运输大数据应用案例

2.1 基于大数据的交通拥堵治理

随着中国经济社会的发展和人民生活水平的提高,汽车保有量快速增长,部分人口密集型城市的人均行车面积不足0.03 km²,城市交通拥堵越来越严重,由此引起的出行时间增加、燃油浪费、尾气污染也愈加严重,社会公众的抱怨也越来越多。为解决交通拥堵问题,不少城市限量发放车辆牌照,限制汽车通行,引起

社会反响且矛盾急剧恶化。

在无法多修路以增加路网密度的情况下,利用大数据分析预测并治理拥堵不失为一条有效途径。通过交通外围感知设备实时获取道路占有率、流量、速度、饱和度等指标,及时感知道路拥堵指数;运用交通专业模型及算法,在地理信息空间上进行相关路网承载能力计算,判定拥堵情况是常态化拥堵还是临时性瓶颈拥堵,找出交通组织的差距;然后对相关的路网及路口进行分析及关联,找出拥堵问题的解决策略。产生拥堵的因素很多,在大数据分析的基础上,找出造成拥堵的关键因素,如外地车流量及货车通行比例过高,可通过制定限行策略进行限流;如遇路口通行不畅,可以通过改变信号机配时方案,提升路口的通行能力;如路网局部承载的交通流过高时,可以通过实时智能分流策略,进行诱导与信号同步,有效引导交通流向^[9-10]。运用大数据进行交通拥堵治理的技术框架如图4所示。

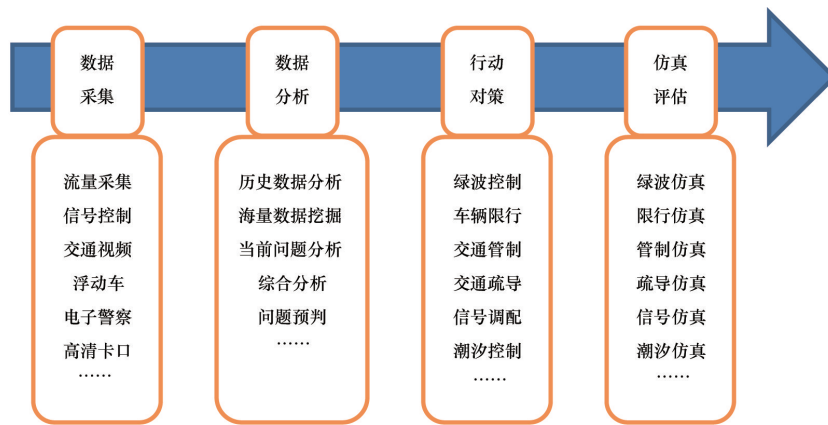


图4 运用大数据进行交通拥堵治理的技术框架

Fig. 4 Technical framework for using big data in traffic congestion management

常发性交通拥堵,可通过仿真软件和智能算法,对路网及车流量的关系模型进行匹配,预测和推演大型交通流的生成、造成交通拥堵的时间节点、交通拥堵的形成和疏散位置等;通过仿真计算,预测下一个周期或者下一阶段的交通拥堵行为,为交通疏导和交通控制提供辅助决策建议。

基于实时大数据分析的方法,解决了在交通拥堵的发现、拥堵形成机理、拥堵调控策略、拥堵疏散影响分析、交通流恢复环节严重脱节的问题,可取得良好的、可操作实施的有效解决拥堵方案。

2.2 基于大数据的高速公路智能稽查系统

中国高速公路的快速发展,尤其是电子不停车收费(electronic toll collection, ETC)的应用,给群众出行带来了便捷。但是电子不停车收费全国联网后,往往存在单一站点通行费数额巨大的现象,在巨大利益的

驱动下,高速公路偷逃通行费的问题也日益突出,部分车主采用倒卡、换卡、假牌、套牌、大车小标等违法手段,以逃避缴纳高速通行费。过去高速公路收费稽查主要依赖于人工,工作量大而且效率很低^[8]。

采用基于大数据分析及数据挖掘技术的智能稽查(图5),主要分析高速公路收费的交易流水数据(包括结构化数据和非结构化数据)及客户服务系统数据,运用多元异构数据预处理技术及聚类、决策树等数据挖掘算法,自动形成节假日免费、ETC流水、特情车辆、绿色通道车辆、异常车辆等的可疑名单,帮助稽查人员稽查^[11-14]。根据数据的来源及产生范围,可以在高速公路省中心、路段分中心、站级收费点分别应用,提高了高速公路收费稽查的效率和查得率,有效打击了日益猖獗的高速公路偷逃通行费行为。



图5 基于大数据的逃费稽查体系

Fig. 5 Big data-based escape auditing system

2.3 基于手机大数据的客流统计

交通出行调查是上至制定交通出行规划,下至公交车辆排班的基础。在过去,交通费调查每年都需花费大量的人力、物力和财力。因此,利用手机在移动通信网络中的定位信息来分析推算动态客流状态,作为一种广域动态客流探测技术,不但可以节省经费,还可获得实时的客流数据。利用时间序列的手机数据分析,可以得到手机用户的出行链信息^[15-17]。

系统将手机的所有信令数据匹配到道路交通网络上,即获得手机用户在道路网络中位置的电子脚印(图6)。在手机数据的准备过程中,可采用寻找次级出行链并通过定义空间阈值和时间阈值的方法,识别出用户停留位置。分析网络中各手机用户出行行为,可以持续把握出行者位置变化信息的特征,进而推断出出行者的出行起讫点、出行时间、使用什么交通工具等出行特征数据。

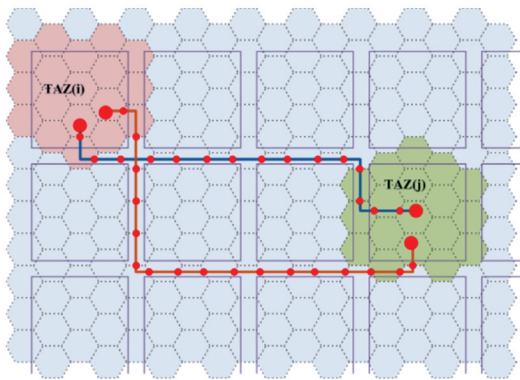


图6 移动通信网络中手机用户出行轨迹
Fig. 6 User travel track of mobile phone in the mobile communication network

一般情况下,基于手机大数据的客流统计采集的多是部分样本的数据,需将大比例有效抽样数据扩样至总体人群的出行数据,总体扩样的平均绝对误差可控制在15%以内。

2.4 基于大数据的民航空管

随着中国航空运输量的不断增加,飞行流量也在增加,而保障能力和空域容量并没有按比例扩展,繁忙机场和空域航班的延误日益严重。近年来航班延误普遍发生,旅客满意度下降,长时间航班延误引起的社会负面影响事件频发。

基于大数据的民航空管系统,通过融合处理海量多源数据(例如空域数据、管制规定与协议等静态信

息,飞行计划、动态报文、实时雷达航迹等动态信息),以各个航路计算结点为基础,利用分布式文件系统和分布式数据库,进行分布式部署,采取大规模并行计算,利用四维(4D)航迹预测航空器飞行航迹中各点空间位置、飞行流等,对机场及繁忙航路、航段及进离港点在的关键空域单元运行容量进行评估(图7),动态推算航班协同放行时间^[18-19]。

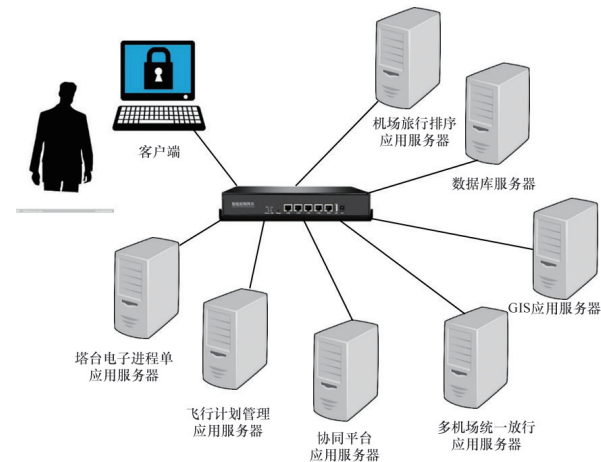


图7 空管流量管理及机场决策协同平台系统拓扑图
Fig. 7 Topology diagram of air traffic management and airport decision coordination platform system

基于大数据的民航空管系统建成以来,在增强空管决策能力、充分利用空域与机场资源、提高航空运营效益、改善服务质量等方面赢得社会各界的广泛认可,保障了航班运行关键时刻、停机位时刻、灵活使用空域、流量控制等重要信息的获取和高效传输,推动了航空运输量和服务质量的提升。

3 交通运输大数据发展及应用趋势

交通运输大数据的应用,尤其是建立在数据共享基础上的数据分析和应用,为交通运输规划、管理、公众出行服务提供了更多、更科学、更实时的依据。随着各类与交通运输相关的数据采集设施设备的丰富,以及数据“烟囱”的消除,交通大数据的应用将出现以下的趋势。

1) 数据分析不再基于随机或伪随机样本,而是基于全体数据。过去由于交通运输信息的采集、存储和处理手段都有限,为让相关的分析变得简单,一般会把数据量缩减到最小,采用常用的统计学原理,利用抽样

数据得出分析结论,从最少的数据里获得最多的信息。而随着采集数据量的增加,以及云存储、云计算等技术的发展,交通运输数据渐渐由少量数据样本分析过渡到大样本分析,最后将发展到全样本的数据分析^[20]。

2) 交通运输大数据混杂性更强,处理结果不一定会更精确。据不完全统计,交通运输数据中的结构化数据只占约5%左右,剩下的都是视频、图像等非结构化数据,而采集到的多源数据可能还会相互矛盾。但随着数据量变得足够大,不精确性对整个分析结论的影响力变小。

3) 适用的数据分析算法不同。数据量较小时,简单算法运算结果表现差,复杂算法运算得到的结果比较准确;而当数据量扩大后,往往会出现简单算法推算出的结果更准确、复杂算法计算出来的结果反而不如简单算法结果准确的现象。

4 结论

综合交通出行大数据开放云平台,为交通运输大数据在交通拥堵治理、高速公路智能稽查、客流统计及民航空管等方面的应用提供了一定的条件,并为大数据分析在交通运输领域更大范围内应用打下了基础。

大数据在交通运输领域中的应用,会使人们重新认识交通需求以及交通运行的内在规律和建模。以大数据为基础进行分析和预测,需要对交通运输系统的规律重新建模,大数据的应用本身(如导航、交通信息服务等)也会改变交通运输的规律。为推进交通运输领域大数据的应用,需建立数据定义、处理、交换等标准的完整体系,各级交通运输主管部门间、政府与企业间需密切合作,才能在大数据应用于交通运输领域中取得共赢。

参考文献(References)

- [1] 郭伟伊,王立伟. 浅谈新型城镇化交通大数据的研究与分析[J]. 科技经济市场, 2017(11): 21-25.
Guo Weiyi, Wang Liwei. The research and analysis of new urbanization traffic big data[J]. Science and Technology Economic Market, 2017(11): 21-25.
- [2] 中国电子技术标准化研究院. 大数据标准化白皮书[R]. 北京: 中国电子技术研究院, 2014.
China Electronic Technology Standardization Institute. Big data standardization white paper[R]. Beijing: China Electronics Technology Research Institute, 2014.
- [3] 保丽霞,王秋兰,沈明,等. 城市智能交通大数据平台的模型库研究与设计[J]. 交通与运输(学术版), 2017(2): 28-31.
Bao Lixia, Wang Qiulan, Shen Ming, et al. Model base study and design for urban intelligent transport big data platform[J]. Traffic & Transportation, 2017(2): 28-31.
- [4] 李丽凯. 苍穹之下: 滴滴快的的“大数据”蓝图[J]. 上海信息化, 2015(9): 60-62.
Li Likai. Under the sky: Big data blueprint of Didi Kuaidi[J]. Shanghai Information Technology, 2015(9): 60-62.
- [5] 段闵. 基于政企合作模式的交通出行服务与管理研究[J]. 中国管理信息化, 2015(17): 162-164.
Duan Min. Study on traffic service and management based on cooperative mode of government and enterprise[J]. China Management Informatization, 2015(17): 162-164.
- [6] 刘冬梅. 政企握手大数据托起众创出行云[J]. 中国交通信息化, 2016(12): 77-78.
Liu Dongmei. Government and enterprises shook hands in big data hold up the public travel clouds[J]. China ITS Journal, 2016(12): 77-78.
- [7] 庞小培. 基于政企合作模式的吉林省综合交通出行信息共享应用科技示范工程研究[J]. 公路交通科技(应用技术版), 2017(1): 129-134.
Pang Xiaopei. Study on the application of science and technology demonstration project of Jilin Province based on the mode of cooperation between government and enterprise[J]. Journal of Highway and Transportation Research and Development, 2017(1): 129-134.
- [8] 孙加君,张亮,王敬陆. 基于大数据的交通拥堵缓解方法探索[J]. 警察技术, 2015(3): 18-20.
Sun Jiajun, Zhang Liang, Wang Jinglu. Exploration of traffic congestion mitigation methods based on big data[J]. Police Technology, 2015(3): 18-20.
- [9] 孟凡林. 基于浮动车大数据的城市交通拥堵自动辨识与可视化系统[D]. 西安: 长安大学, 2016: 27-34.
Meng Fanlin. Automatic identification and visualization system of urban traffic congestion based on big data of floating vehicle [D]. Xi'an: Chang'an University, 2016: 27-34.
- [10] 赵鹏军,李锐. 大数据方法对于缓解城市交通拥堵的作用的理论分析[J]. 现代城市研究, 2014(10): 25-30.
Zhao Pengjun, Li Kai. Performance analysis of the effect of big data method on urban traffic congestion[J]. Modern Urban Studies, 2014(10): 25-30.
- [11] 张一衡,沈刚. 重点营运车辆动态数据在智慧高速中的应用[J]. 中国交通信息化, 2015(增刊1): 22-24.
Zhang Yiheng, Shen Gang. Application of key commercial ve-

- hicle dynamic data in smart high speed[J]. China ITS Jowmal, 2015 (suppl 1): 22-24.
- [12] 秦振威, 李金峰, 徐铤. 高速公路收费稽查管理系统的设计与实现[J]. 数字技术与应用, 2016(4): 156-156.
Qin Zhenwei, Li Jinfeng, Xu Ting. Design and implement of highway toll inspection management system[J]. Digital Technology and Applications, 2016(4): 156-156.
- [13] 张晓航, 任文龙. 基于数据挖掘的高速公路联网收费稽查研究[J]. 软件, 2011(11): 57-59.
Zhang Xiaohang, Ren Wenlong. Research on highway network charging audit based on data mining[J]. Software, 2011 (11): 57-59.
- [14] 李嘉. 基于数据的高速公路智慧型稽核系统的研究与实现[D]. 成都: 电子科技大学, 2015: 57-61.
Li Jia. Study and implementation of intelligent highway audit system based on data[D]. Chengdu: University of Electronic Science and Technology of China, 2015: 57-61.
- [15] 阳扬, 欧冬秀, 何向俊. 基于手机APP大数据的交通出行数据获取方法[J]. 交通信息与安全, 2015(6): 40-47.
Yang Yang, Ou Dongxiu, He Xiangjun. Data access method of trip based on mobile APP data[J]. Journal of Transportation Information and Security, 2015(6): 40-47.
- [16] 陆俊贤, 裘炜毅, 庞海峰. 基于手机网络数据的实时动态区域集散客流检测研究[C]//中国智能交通年会. 合肥: 中国智能交通协会, 2013: 172-178.
Lu Junxian, Qiu Weiyi, Pang Haifeng. Study on real-time dynamic regional distributed passenger flow detection based on mobile phone network data[C]//China Intelligent Transportation Annual Meeting. HeFei: China Intelligent Transportation Association, 2013: 172-178.
- [17] 胡忠顺, 王进, 朱亮. 基于手机信令数据的大客流监控应用研究[J]. 电信技术, 2017(4): 21-25.
Hu Zhongshun, Wang Jin, Zhu Liang. Application of large passenger flow monitoring based on mobile phone signaling data[J]. Telecommunications Technology, 2017(4): 21-25.
- [18] 高红旭, 康永, 郭芑. 大数据技术在民航空管监控系统中的应用[J]. 现代导航, 2015(2): 144-149.
Gao Hongxu, Kang Yong, Guo Peng. Application of large data technology in air traffic control system of civil aviation[J]. Modern Navigation, 2015(2): 144-149.
- [19] 卜强. 民航空管迎来大数据时代的挑战[J]. 空中交通, 2015 (11): 68-71.
Bu Qiang. Civil aviation air traffic control ushered in the challenges of large data age[J]. Air Traffic, 2015(11): 68-71.
- [20] 维克托·迈尔-舍恩伯格. 大数据时代: 生活-工作与思维的大变革[M]. 杭州: 浙江人民出版社, 2013.
Victor Mayer-Schoenberg. Big data age: A revolution that will transform how we live, work, and think[M]. Hangzhou: Zhejiang People's Publishing House, 2013.

Application of big data of transportation

YANG Qi¹, LIU Dongmei²

1. Research Institute of Highway, Ministry of Transport, Beijing 100088, China
2. National Intelligent Transport Systems Center of Engineering and Technology, Beijing 100088, China

Abstract With the wide applications of information and intelligence in transportation services and management, large and rich data resources are accumulated in the industry. This paper analyzes the characteristics, the types and the data volume of the transportation big data, and briefly describes the structure, the function and the application of the integrated traffic travel big data open cloud platform, and analyzes application cases in several aspects, such as the traffic big data in the traffic congestion control, the highway intelligent inspection, the passenger flow statistics and the civil aviation management. Finally, it is suggested that the transportation data are very hybrid in nature, and the processing results may involve contradictory conclusions. As the data amount increases, the influence of inaccuracy on the whole analysis conclusion becomes smaller. When the data volume is expanded, simple algorithms may be more practical.

Keywords transportation; big data; open cloud platform ●



(责任编辑 王志敏)