

文章编号:1005-0523(2016)02-0056-07

大数据技术在轨道交通应急辅助决策系统中的应用设计

刘峰博,干叶婷,周 峰

(同济大学道路与交通工程教育部重点实验室,上海 201804)

摘要:从城市轨道交通网络化运营条件出发,针对应急辅助决策系统将要应对越来越复杂的异构大数据问题,探讨大数据技术在系统中的应用。首先分析系统中大数据的组成、特征和处理需求;随后基于可借鉴的大数据处理技术,构建了应急辅助决策系统中大数据的处理框架,并设计了数据采集、数据预处理、数据存储、数据分析/挖掘和数据结果展现等5个环节的大数据处理技术应用方法;最后基于北京市城市轨道交通现行方案,设计了融合大数据技术的应急辅助决策系统流程。研究可为城市轨道交通应急管理提供技术参考。

关键词:轨道交通;应急管理;大数据;辅助决策

中图分类号:U231.92 **文献标志码:**A

DOI:10.16749/j.cnki.jecjtu.2016.02.009

在城市轨道交通运营进入网络化的当下,运营安全与可靠性是至关重要的问题。轨道交通运营管理者需要对日常的运营质量、治安事件加以控制,对大客流冲击、突发事件等高效应对,以保障运营服务水平。应急辅助决策系统负责开展网络运营数据采集发布、参数模式管理、系统维护,以及对运营数据、业务指标等进行分析,为运营和建设提供辅助决策依据。需要轨道交通运营各专业系统及地面公共、消防、公安等外部资源的支持,需要收集和分析跨线路、跨专业、跨部门、跨行业的信息,以满足网络应急事件的快速响应、高效处置。

以往关于城市轨道交通应急管理的研究,数据集成、信息处理往往是技术关注要点。邵伟中^[1]、徐瑞华^[2]等人研究了网络应急处置辅助决策的主要技术及相关问题,在基于工作流的轨道交通应急管理系统设计与实现中体现了贯穿始终的是信息处理,数据的频繁下载上传^[3]。于鸿飞^[4]等人在阐述应急预案和案例的检索技术时充分说明了数据的多维性,用案例推理和专家知识方法分类数据,不同类型数据用不同算法计算。李文杰^[5]等人针对目前已有的城市交通安全运营、应急管理的信息系统,设计了基于SOA的应急管理系统集成方案,实现了应急管理各类信息的共享和业务流程的协同处理。Yi Peng^[6]等构建了一个基于数据集成、数据挖掘和多目标决策支持的突发事件应急信息管理框架。Vagelis Hristidis^[7]等系统详细地讨论了突发事件应急管理每个过程中的数据管理及分析问题,包含多参数实时动态信息获取、网络监测、海量数据集成、特征参数识别、网络信息传播、信息交互共享等。应急辅助决策系统的基础和关键是实时的数据处理和信息传输。目前系统内部的数据处理主要采用传统的关系数据库方法,要求数据的一致性和关联性;系统与运营各专业系统和外部资源间的信息沟通则主要通过专用的通信设施进行,要求高的稳定性和可靠性。

在网络化运营越来越复杂的情况下,城市轨道交通应急辅助决策系统要应对越来越复杂的数据采集、数据处理、数据挖掘和数据发布等需求,数据大量、连续、分布式、异构、不断变化,且应急决策要求即时响

收稿日期:2015-07-27

基金项目:中国博士后科学基金(2014M551454)

作者简介:刘峰博(1990—),女,博士研究生,研究方向为交通运输规划与管理。

通讯作者:周峰(1986—),男,博士,研究方向为轨道交通运营管理。

应,传统的关系数据管理(并行数据库)和不同部门系统间的互通方式均存在扩展性和适应性不足的问题,将无法胜任数据分析越来越“大、多、快、准”的高要求。而这些复杂的数据正符合大数据的特征,因此处理大数据可参考当前比较先进的技术,在数据分析和信息传输方面寻找更可靠和长远的方式。借鉴道路交通领域应用大数据技术的经验^[8-11],探讨大数据技术在轨道交通应急辅助决策系统中的应用。

1 应急辅助大数据

1.1 主要数据组成

1) 客流监视数据。网络化运营时代的城市轨道交通,对客流的预测和监视是从网络、线路和车站3个层面出发。针对特定日期,比如国定节假日进行客流预测,以调整和实施新的运行图,并为大客流下的一些重要换乘站、一些繁忙线路以及整个网络层面的运营协调提供预警,为大客流应急指挥提供辅助支持。与此同时,在运营时段进行实时的客流监视,以保障在突发情况下能够及时而灵活地应对。客流数据是基础,需分层分布采集和处理。

2) 列车监视数据。对运营中的列车和热备列车进行实时的追踪检测:一方面能掌握其位置,确保其正常工作;另一方面辅助应急情况的应对策略,以求在突发事件后能够在第一时间确定列车位置,实施列车清客、列车救援等行车调整方案。列车监视数据包括对整个网络中每一列车的每一项状态指标的实时监视。

3) 设备监视数据。城市轨道交通应急辅助决策同时也需要设施设备的实时数据,比如信号系统设备是否正常运转、通信设备是否通畅等,这些数据是应急情况下下达列车运行调整策略、客运组织调整计划等的基础保障。

4) 电子化预案数据。为了便于机动化的应急调度指挥,目前北京、上海等城市轨道交通网络发达的城市均采用电子化、数字化的应急预案系统。将设定好的预案输入到预案管理系统中,应急情况下结合调度规则和事故案例经验,能够快速调用系统中适合当下情况的预案。对预案数据进行数字化管理、调用分析与调度指挥相适应。

5) 应急处置数据。主要指应急事件处置报告、图片、视频、音频等数据,是针对具体事件的原始资料和处置结果,此类数据是应急处理和案例学习的重要参考,通常包括很多细节,数据量大且不具有相对统一的格式。

1.2 数据特征

1) 一般性特征。应急辅助数据是上述多个系统和环节的所有数据集,是轨道交通的大数据,具有“4V”特征。规模性(Volume):数据的规模从MB级别向TB跃进,甚至将来可能是PB级;多样性(Variety):除了传统的关系型数据,视频、音频和图片等非关系型数据量也越来越大;高速性(Velocity):应急辅助大数据处理对时效性的要求非常高;价值性(Value):从海量数据中得到支持决策的信息是一个很大挑战。如何对TB、PB级别的轨道交通数据进行准确、高效地处理和分析及预测,做出即时和正确的应急处置决策以保障城市轨道交通正常运营是目前亟待解决的关键问题。

2) 专业性特征。轨道交通应急辅助数据同时具有其独有的一些特征。一方面,客流数据、列车数据、设备数据等,它们的核心内容都具有结构化特征,有其各自统一的格式,具有可通用的数据模型、数据标准、指标体系等。另一方面,预案数据、应急处置数据等较难形成统一的结构,目前大多以文件个数来统计。数据的不同结构特征为建立完善即时的应急辅助决策系统带来了大挑战。

图1列举了目前北京城市轨道交通网络化运营中应急决策所需的相关数据,可见数据量大、来源丰富且数据类型多样。根据符合实际的估算,结构化数据约50TB,非结构化数据约86TB,非结构化数据比例超过60%。

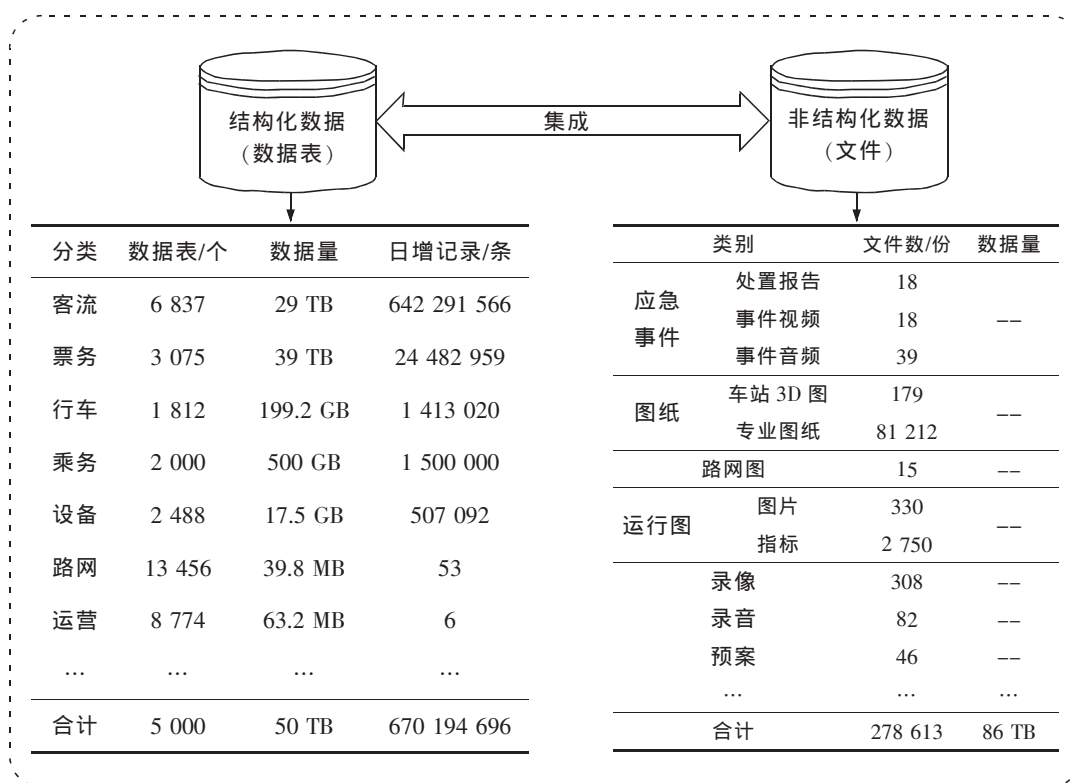


图1 城市轨道交通大数据举例(作者修订源于文献[12])

Fig.1 Examples of urban rail transit big data

1.3 数据处理需求

城市轨道交通网络化环境下,对大量化、异构化数据处理的高效性和可用性要求较高,在应急条件下,尤其需要快速地提取到有效的预案信息、行车数据、客流数据等,同时要第一时间记录事件的录像、音频等资料,以及后期的处置报告、案例分析等的存档与管理。

传统的并行数据库技术无法胜任这种异构化的数据,因为它依靠并行计算提升数据处理速度,追求高度一致性和容错性。根据 CAP (一致性 Consistency, 可用性 Availability, 分区容错性 tolerance to network Partitions)理论,在分布式系统中,三者不可兼得,因而并行关系数据库必然无法保证较好的可用性和扩展性^[13]。因此,必须寻找高扩展性的数据分析技术,在数据的采集、预处理、存储、分析与挖掘、成果展现的全过程中有效应用,为应急决策提供辅助,大数据的处理技术为此开拓了新思路。

2 大数据技术的应用

2.1 可借鉴的大数据处理技术

大数据处理技术主要指基于 MapReduce 思想实现的并行计算技术,从基本的 MapReduce 平台到 Hadoop 平台、Spark 平台,数据的处理效率越来越高。轨道交通应急管理在强调快速响应的同时,也较为关注任务管理和数据计算的分布明确性,而 Spark 是基于内存迭代计算的快速数据分析解决框架,不具备对计算任务的管理能力(需要第三方资源管理平台调度分配任务)^[14],因此暂不考虑 Spark 处理技术。

1) MapReduce。MapReduce 技术框架包含 3 个层面的内容:① 分布式文件系统;② 并行编程模型;③ 并行执行引擎。文件系统数据存储采用键/值对(Key/Value)模式,元数据集中管理与数据块分散存储,通过数据复制(至少 3 份)实现高度容错。并行编程模型通过 Map(分解)和 Reduce(合并)2 个阶段,先分后合地按照 Key/Value 对处理数据。并行执行引擎完成了这一系列并行计算过程。

该技术将传统的查询、分解及数据分析进行分布式处理,将处理任务分配到不同的处理节点,因此具有更强的并行处理能力。其对数据一致性要求不高,具有突出的扩展性和可用性,特别适用于海量的结构化、半结构化及非结构化数据的混合处理。通过结合 MapReduce 技术进行实时分析,某家电公司的信用计算时间从 33 h 缩短到 8 s,而 MKI 的基因分析时间从数天缩短到 20 min^[15]。图 2 是比较典型的基于 MapReduce 技术的大数据处理工具。

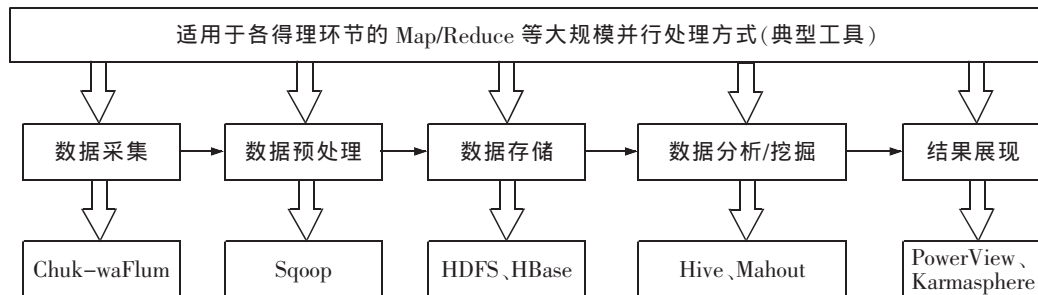


图 2 基于 MapReduce 技术的大数据处理工具

Fig.2 Big data processing tools based on MapReduce

2) Hadoop。单纯的 MapReduce 作为典型的离线计算框架,无法满足在线实时计算需求。Hadoop 是开放源码并行运算编程工具和分散式档案系统,是 MapReduce 的开源实现,凭借其开源和易用的特性,成为大数据处理的首选^[16]。系统存储由 HDFS(分布式文件系统)实现,分析由 MapReduce 实现。Hadoop 架构支持在公有云端存储 EB 量级数据的应用,许多互联网公司,包括 Facebook、谷歌、eBay 和雅虎等,都已开发了基于 Hadoop 的 EB 量级超大规模数据应用。

国内外大数据技术的发展不断创新,一些大集团更是在争先推出适应自身的大数据处理技术和系统。因此,要从大体量、多类别、异构化的数据中快速提取价值,需要结合以 MapReduce 和 Hadoop 为代表的非关系数据分析技术,改进或重构整个数据库技术体系。2011 年,Facebook 首次公开新数据处理分析平台 PUMA,通过对数据多处理环节区分优化,相比之前单纯采用 Hadoop 和 Hive 进行处理的技术,数据分析周期从 2 d 降到 10 s 以内,效率提高数万倍^[16]。

2.2 大数据处理技术的应用设计

轨道交通的大数据分析不能简单直接地使用一款数据分析软件,而是要在数据处理需求的基础上,分析如何实用到各环节中。通过对轨道交通突发事件应急管理数据集成、数据挖掘及决策支持的分析,首先构建了应急辅助决策系统的大数据信息处理框架,如图 3 所示;其次分析大数据处理技术在各环节中的应用。

1) 数据采集。对于城市轨道交通运营组织数据这类保密性要求较高、涉及部门或系统较广的数据,应该使用分布式异构数据接口、特定 API(应用程序编程接口)采集,并传到统一数据接口共享特定数据。对于一些非结构化数据,建立统一的本地数据文件,以结构化的形式存储。在发现采集的数据与规定的标准或规范不符时,应及时修正,为应急情况下的即时响应做好数据基础准备。

2) 数据预处理。数据预处理即提取有效信息。现实数据经常存在不完整或不一致、数据存在冗余的情况,需要对应急处置决策信息数据进行填补、筛选、合并、规格化、检查一致性等预处理,并对数据的多种属性进行初步组织,从而提高数据分析质量,减少分析时间,可以利用基于 Hadoop 的数据处理技术,以提高处理效率。

3) 数据存储。目前数据的存储采用很多台 PC 存储,那么就需要保证其可用性,因为一旦发生错误,数据失效,将给正常运营和质量分析带来负面影响。可应用 HDFS 采用 Key/Value 的存储模式,元数据集中管理与数据块分散存储,且便于调用。另外,建议对重要的同一份运营数据在不同节点上存储 3 份副本,以提高系统的容错性和可靠性,为应急辅助决策提供有效的数据支持。同时,整个应急处置过程中的数据要与运营公司运营数据做到同步,双方不断更新实时应急数据和运营数据,为线网运营提供动态保障。

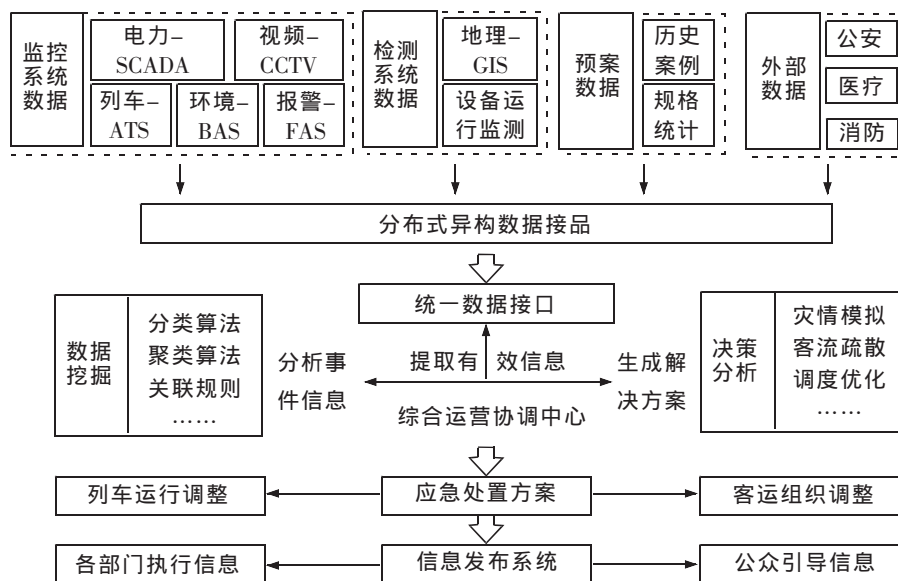


图3 应急辅助决策系统的大数据处理框架

Fig.3 Big data processing frame of emergency decision supporting system

4) 数据分析/挖掘。对提取到的有效信息进行分析挖掘得以生成解决方案,或者对应急处置的资料数据、决策信息、处置结果进行必要的分析研究时,可以采用并行处理方式以提高异构数据的处理效率,如上文提到的 MapReduce 技术。在处理过程中,各处理节点完成自身任务后,只需将结果返回给主控制节点即可,各处理节点间不进行任何交互;并且当数据量增加时,要求只增加分布式服务节点,无须修改分析算法。

5) 数据结果展现。大数据场景下,结果展现更加注重交互式 and 可视化,主要形式有报表、图形化、选择性查询等。在综合运营协调中心、运营公司等用户使用应急辅助决策系统时,为了快速方便地获得自己所需的信息,需要与系统进行交互,如:对各类查询条件进行组合筛选、增删查询条件、界面布局个性化定制等,从而增加信息获取的灵活性和便捷性。

3 大数据技术的实践设计

基于北京市城市轨道交通应急管理的方法,设计了融合大数据技术的应急辅助决策系统的业务流程,如图4所示。其中白色箭头均代表数据的传递和处理,存在于应急处置的各个环节。数据同步意味着突发事件时的运营相关数据(客流、票务、行车、乘务、设备、路网、运营等结构化数据)与应急事件数据(事件信息、应急预案、录音录像、路网图、运行图等非结构化数据)的采集、预处理、存储、分析挖掘、结果展现等环节的快速、统一处理,且实现各个指挥调度终端的同步跟进。

“自动报警”、“人工报警”主要包括数据的采集阶段,由北京市 TCC(轨道交通路网指挥中心)指挥系统根据实时数据自动报警或者运营公司实地运营数据发生异常后由调度相关人员报警;随后应急指挥中心对突发事件接警,进行前期数据记录和分析,由此利用人机交互界面选择适当的预案执行,各级各部分对执行过程的进展进行实时监控。这一突发事件应急处置的基本流程中,应急指挥中心和运营公司在每一环节都要做到数据同步更新,以达到实时掌握运营动态,向调度人员、列车司机、乘客等发布实时有效的信息。决策信息最终以各类形式显示,调度中心和运营公司再进行处置操作的评估报告分析,得到的数据报表可以存入应急预案系统,以备日后工作参考。

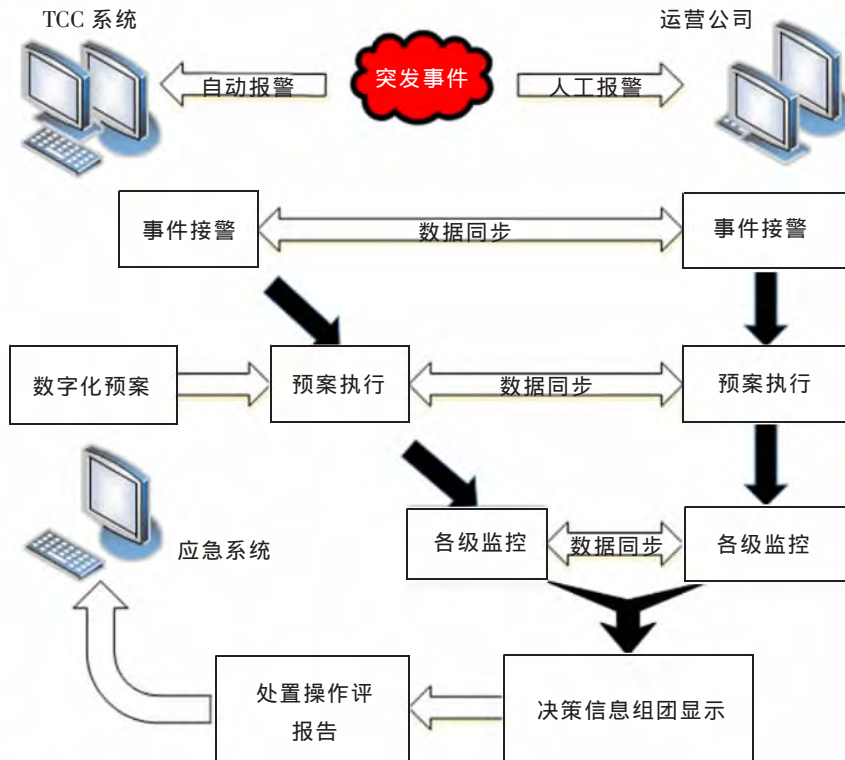


图 4 大数据技术的应急辅助决策系统业务流程设计(修改于文献[12])

Fig.4 Business procedure of emergency decision supporting system based on big data technologies

4 总结与展望

在分析城市轨道交通应急辅助大数据组成和特征的基础之上,探讨了城市轨道交通应急辅助决策系统中大数据方法的应用框架及关键技术。综合上述分析可见,应急辅助决策系统对数据的要求符合大数据的相应特征,其对大数据技术有较大的需求,应用前景广阔。在轨道交通大数据技术的实践中,有以下 3 点建议:

1) 大数据技术中不是单纯地套用现有技术,而是要结合轨道交通和应急处置的具体需求,进行实用性应用的开发和实践。

2) 数据是基础和中心,也是一切大数据应用的前提,因此需要先做好数据准备: 建立数据标准体系,确保数据质量; 采集尽可能多且全面的业务数据; 运用数据模型对数据进行整合,建立数据联系; 建立数据安全管理体系,确保数据安全。做好规范的数据管理,为数据分析等后续工作提供可靠依据。

3) 在对数据管理整合之后,通过各个业务层进行循序渐进地数据分析应用和推广,最终达到为安全可靠地运营工作提供保障的目的。

参考文献:

- [1] 邵伟中,徐瑞华. 城市轨道交通网络运营协调及应急处置辅助决策技术[J]. 城市轨道交通研究,2008,11(6):17-18.
- [2] 徐瑞华,滕靖. 城市轨道交通网络应急指挥辅助决策支持系统的相关问题研究[J]. 城市公用事业,2010,24(3):1-4.
- [3] 徐瑞华,张铭,王志强. 基于工作流的轨道交通应急管理系统设计与实现[J]. 同济大学学报:自然科学版,2008,36(6):754-759.
- [4] 于鸿飞,秦勇,王子洋,等. 城市轨道交通应急处置辅助决策系统的研究[J]. 交通信息与安全,2013,31(5):163-168.

- [5] 李文杰,秦勇,贾利民. 基于 SOA 的城市轨道交通应急管理系统集成的研究[C]//中国职业安全健康协会.第一届全国安全科学理论研讨会论文集.北京:中国商务出版社,2007:367-375.
- [6] PENG Y,ZHANG Y,TANG Y,et al. An incident information management framework based on data integration, data mining, and multi-criteria decision making[J]. Decision Support Systems,2011,51(2):316-327.
- [7] HRISTIDIS V,CHEN S C,LI TAO,et al. Survey of data management and analysis in disaster situations[J]. Journal of Systems & Software,2010,83(10):1701 - 1714.
- [8] 周为钢,杨良怀,龚卫华,等. 大数据处理技术在智能交通中的应用[C]//中国智能交通协会.第八届中国智能交通年会论文集.北京:电子工业出版社,2013:226-235.
- [9] 叶亮. “大数据”背景下我国交通数据管理应用的转型与发展[J]. 交通与运输:学术版,2013(S2):65-68.
- [10] 邱卫云. 智能交通大数据分析云平台技术[J]. 中国交通信息化,2013(10):106-110.
- [11] 唐要安. 大数据在交通中的应用[J]. 交通世界:运输车辆,2013(12):126-127.
- [12] 北京市轨道交通指挥中心, Teradata Universe. 首都轨道交通大数据实践与探索[E/B/OL]. 百度文库, <http://dwz.cn/1Hzea1>.
- [13] 李国杰,程学旗. 大数据研究:未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考[J]. 中国科学院院刊,2012,27(6):647-657.
- [14] 梁彦. 基于分布式平台 Spark 和 YARN 的数据挖掘算法的并行化研究[D]. 广州:中山大学,2014.
- [15] 李志刚,朱志军,余从国,等. 大数据:大价值、大机遇、大变革[M]. 北京:电子工业出版社,2012:152-178.
- [16] 陶雪娇,胡晓峰,刘洋. 大数据研究综述[J]. 系统仿真学报,2013,25(S1):142-146.
- [17] 孟小峰,慈祥. 大数据管理:概念、技术与挑战[J]. 计算机研究与发展,2013,50(1):146-169.
- [18] 王树良,丁刚毅,钟鸣. 大数据下的空间数据挖掘思考[J]. 中国电子科学研究院学报,2013,8(1):8-17.
- [19] 覃雄派,王会举,杜小勇,等. 大数据分析——RDBMS 与 MapReduce 的竞争与共生[J]. 软件学报,2012,23(1):32-45.
- [20] 李伟,徐瑞华. 突发事件下地铁网络乘客出行行为仿真模型[J]. 华东交通大学学报,2015,32(2):46-53.

Application Design of Big Data Technologies in Emergency Decision Supporting System

Liu Fengbo, Gan Yeting, Zhou Feng

(Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China)

Abstract: According to operating conditions of urban rail transit network, aiming at the issue of more and more complex heterogeneous big data that the emergency decision supporting system would deal with, this paper discusses the application of big data technologies in the emergency decision supporting system. Firstly, the composition, characteristics and processing requirements of big data are analyzed. Secondly, in light of the referential processing technologies, the processing framework of emergency decision supporting system is established. Besides, the technology applications of big data processing methods in data collection, preprocessing, storage, analysis/mining and data results display are carried out. Finally, based on the present urban rail transit programs of Beijing, procedures of emergency decision supporting system integrating big data technologies are designed. The study can provide technical references for urban rail transit emergency management.

Key words: urban rail transit; emergency management; big data; decision supporting system

(责任编辑 姜红贵)