

交通大数据及应用技术研究

于 硕 李泽宇

(大连财经学院工商管理学院, 辽宁 大连 116622)

摘要: 随着目前计算机科学技术的飞速发展以及计算机硬件对数据处理能力的飞速提升, 人类已经进入了数据信息时代。文章从宏观方面对大数据在城市交通中的应用进行了阐述, 并说明了现今交通大数据的挑战以及目前对大数据的几种常用存储方案以及常用的处理技术。

关键词: 交通大数据; 大数据存储方案; 大数据处理技术; 数据信息 **文献标识码:** A

中图分类号: U495 **文章编号:** 1009-2374 (2017) 04-0090-02 **DOI:** 10.13535/j.cnki.11-4406/n.2017.04.046

1 大数据概述

1.1 大数据定义

大数据, 又称海量数据、巨量数据等, 通常指那些无法通过常规存储媒介存储, 无法通过人工或者常规数据处理软件进行管理和处理的数据集合。大数据是一个抽象的概念, 究竟多大的数据量可以定义为大数据至今没有一个定论。大数据具有海量的数据规模、快速的数据流转、多样的数据类型和价值密度低四大特征, 也就是说, 大数据不仅仅体现在存储数量级的大小上, 而且还在于数据的获取、存储、维护、处理等方面。

1.2 大数据特点

IBM公司针对大数据的特性提出了大数据的5V特点: Volume (大量)、Velocity (高速)、Variety (多样性)、Value (价值) 和Veracity (真实性)。

其中Volume (大量) 指大数据的数据量级巨大, 可以从几百TB到几百PB甚至EB级别的规模。与几十年前印刷品占据主流的时候不同, 信息时代数据的产生速度非常快, 据国际数据公司 (IDC) 的统计, 仅在2011年时, 全世界范围内被创建和复制的数据总量就有1.8ZB左右, 已经远远超过有史以来人类通过印刷品传播的信息量的总和。据该公司预测, 中国的大数据市场将在2012~2016年间增长5倍, 中国已经真正步入了大数据时代。

Velocity (高速) 指大多数场景下, 我们都要求大数据要在一定的时间限制内得到及时的处理。例如美团、大众点评等团购软件, 通过对用户的地理位置定位实时向用户推荐附近美食、团购优惠等。

Variety (多样性) 指大数据包含各式各样形态以及各种格式的数据源。针对不同场景, 数据有多种存储方式, 例如json格式、xml文件、数据库文件等方式。

Value (价值) 指大数据中蕴藏的有价值的信息, 通过大数据的分析挖掘等手段, 可以带来巨大的商业或学术价值。例如淘宝通过对用户购物以及浏览数据进行挖掘, 从而给用户推荐合适的商品, 为淘宝创造了巨大的利润。

Veracity (真实性) 指大数据的数据来源, 处理结

果要保证与真实情况相符合, 即它们可以如实反映或者折射出社会现实等。例如网上购物网站中, 卖家的刷单情况就会影响数据分析结果对它的正确评价。

2 交通大数据

2.1 概述

随着世界范围内城乡发展的步伐不断加快, 公路网、铁路网的大规模覆盖, 城市的交通建设成为了越来越重要的课题。传统模式下的城市交通建设、管理等在今天出现了种种弊端。由此, 在大数据时代的背景下, 通过大数据的分析、处理, 对城市交通进行动态的指导规划成为了目前大数据的主要现实应用场景之一。

在高德地图的“2014年中国主要城市拥堵分析”报告中, 北京、上海、广州、深圳、哈尔滨和杭州位于前列, 高峰期行车平均速度在20km/h左右徘徊。尤其节假日黄金周期间, 拥堵情况更加严重, 使原本脆弱的城市路网的运行可靠性越来越低。

2.2 交通大数据来源

交通大数据来源广泛, 城市交通网每天产生各类交通数据不计其数。按照类型, 交通大数据的来源大致可以分为来源于人的数据、来源于交通工具的数据、来源于行业监管的数据和来源于其他方面的数据几类。以下将分别就这几类数据来源做简要说明:

2.2.1 来源于人的数据。人作为道路交通的主要参与者, 汽车驾驶员的驾驶轨迹、交通警察所在路口的位置、公交车乘客的刷卡记录以及行人的行走轨迹等都是交通参与者的数据来源。目前, 基于公共交通的电子收费数据已经有了各种丰硕的研究成果和工作。针对该类数据的工作也是目前学术界交通大数据的研究重点之一。

2.2.2 来源于交通工具的数据。车是城市交通中的主体, 按照功能群体可以大致分为公交车、私家车、出租车和网约车4类。每种类型的车都有各自的特点: 公交车路线固定, 而且能够直接获得乘客的刷卡数据, 得到乘客的上下车地点等信息; 私家车涉及的数据相对来说难以获取, 该类数据涉及到驾驶人的个人隐私, 但同时

也是最为有价值的一类数据；出租车没有固定路线，但是却有相对固定的活动范围；网约车是最近几年兴起的一类新兴群体，通过应用平台的私家车主可以利用空闲时间提供服务，缓解高峰期打车难等问题。

此外，基于物联网、车联网等数字终端设备的传感器采集的车辆信息也是交通大数据的重要组成部分。

2.2.3 来源于行业的监管数据。很多特殊行业，如危险品运输、快递服务业、货运公司等行业都会对相应的业务有一定的监管数据，该类数据虽然有一定局限性，但是数据准确性、可靠性非常高。

2.2.4 来源于其他方面的数据。除了上述提到的两大类数据之外，城市道路本身以及天气等客观因素其实也属于交通大数据的一部分。城市各个主干分支路线的拥堵程度以及天气对路况的影响会在一定程度上决定该段道路的拥堵程度，如果能够及时掌握这类数据，通过适当的算法分析以及调配，可以在很大程度上缓解交通压力。

2.3 研究现状及现阶段问题与挑战

目前交通大数据的应用场景越来越多，包括在交通管理中的应用、在智能交通中的应用、在拥堵检测中的应用以及在事故分析与处理中的应用等多种应用场景。

首先，根据交通大数据的分析，可以为城市交通的管理、决策、规划和运营提供有效支持，通过对客流量分析优化交通规划实现按需控制；其次，根据道路流量数据，通过分析预测等手段，随时针对道路突发状况进行处理，比如封闭事故路段，自动将车辆分流至其他路段等；再次，通长时间对交通数据的分析处理，可以找出当前城市交通网中诸如信号灯、交通岗、人行横道等设置的不合理之处加以调整；最后，通过交通大数据我们还可以在提升城市交通网络流畅性的同时，为车联网应用、社会管理等提供一种新的手段和方式。

3 交通大数据存储方案

3.1 MPP关系型数据库

MPP架构的新型数据库集群采用SharedNothing结构，通过粗粒度索引、列存储等技术，结合MPP架构的分布式计算模式完成大数据的分析和存储。该类架构具有高性能和高扩展性的特点，在企业分析类应用领域获得了广泛应用。而且该架构的运行环境多为低成本的服务

3.2 基于Hadoop的非关系型数据库

Hadoop是由Apache基金会开发的分布式系统基础架构，它的核心思想是HDFS和Map Reduce。通过对该项开源技术的扩展和封装，可以将海量数据部署到价格低廉的硬件上，并能够提高其吞吐量。尤其对于处理非结构化数据、半结构化数据以及复杂的ETL流程等，该类型数据库有着独特的优势。

3.3 大数据一体机

大数据一体机是一种专门为大数据设计的软硬件结合产品。它由一组集成的服务器、存储设备、操作系统和数据库管理系统构成。该类存储方式具有很好的稳定性和纵向扩展性。

4 交通大数据处理技术

4.1 Apache Spark

Apache Spark是一套卓越的开源处理引擎，专门面向复杂分析、高速处理和易用性需求而打造。它提供了很多立足于数据结构上的编程接口，因此我们根据需求完成具有特定功能的处理程序。Spark的处理速度极快，而且支持多种语言为其编写应用，能够轻松处理实时数据流，支持复杂的分析操作。

4.2 Hadoop

Hadoop是一个开源的分布式处理软件框架，它的可靠性非常高，它会假设元素存储失败等场景，维护多个数据副本，确保能够对失败的结点做出处理重新分配任务。Hadoop依赖于社区服务器，因此成本低廉。

4.3 MapReduce编程模型

MapReduce是一种针对大规模数据集的并行运算提出的编程模型。它使编程人员在不会分布式编程的情况下能够将程序运行在分布式系统上。它主要运用了映射（Map）和归约（Reduce）的思想，通过指定一个映射函数把一组键值对映射成一组新的键值对，之后制定并发的归约函数，保证所有映射的键值对中的每一个共享相同的键组。

4.4 各类分布式数据库

随着数据量的日益增长，传统的关系型数据库暴露出越来越多的问题，由此，以NoSQL为代表的非关系型数据库开始飞速发展，开始了数据存储从集中式数据库到分布式数据库的转变。分布式数据库直接将数据存储

5 总结与展望

大数据时代已经来临，并且已经应用在现实生活中的方方面面，我们的生活与大数据已经密不可分。交通大数据作为与日常生活息息相关的一个大数据应用领域已经有了很多丰硕成果，但是就应用流程来看，从数据获取到数据存储，从数据处理到结果分析都存在着各种各样的困难与挑战。因此，想要真正能够合理高效地利用大数据还有很长的路要走。

参考文献

- [1] 刘维贵. “大数据”研究综述[J]. 办公自动化: 综合版, 2014, 7(1).
- [2] 陶雪娇, 胡晓峰, 刘洋. 大数据研究综述[J]. 系统仿真学报, 2013, (s1).
- [3] 段宗涛, 郑西彬, 李莹, 等. 道路交通大数据及其关键技术研究[J]. 微电子学与计算机, 2015, (6).
- [4] 陆化普, 孙智源, 屈闻聪. 大数据及其在城市智能交通系统中的应用综述[J]. 交通运输系统工程与信息, 2015, 15(5).
- [5] 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).

作者简介: 于硕(1990-), 女, 吉林长春人, 大连财经学院工商管理学院助教, 硕士, 研究方向: 数据挖掘、图像处理。
(责任编辑: 王 波)