**上海电力大学**

**大数据技术原理及应用课程报告**



题目 智能交通系统中的大数据分析调查

院系 计算机科学与技术专业 电力信息技术

学号 19108002

姓名 崔荣成

2020-06-25

智能交通系统中的大数据分析调查

**摘要**——大数据正成为智能交通系统（ITS）的研究热点，可以在世界各地的许多项目中看到。智能交通系统将产生大量数据。生成的大数据将对智能交通系统的设计和应用产生深远的影响，使ITS更安全，更高效，更有效。在ITS中研究大数据分析是一个很好的领域。本文首先回顾了大数据和智能交通系统的历史和特点。接下来讨论在ITS中进行大数据分析的框架，其中总结了数据源和收集方法，数据分析方法和平台以及大数据分析应用类别。介绍了智能交通系统中大数据分析应用的几个案例研究，包括道路交通事故分析，道路交通流量预测，公共交通服务计划，个人旅行路线规划，铁路运输管理和控制以及资产维护。最后，本文讨论了在ITS中使用大数据分析的一些开放性挑战。

**索引字段**——大数据分析，智能交通系统（ITS），机器学习，运输。

# 一、引言

近年来，大数据已成为科研和行业的热门话题。它代表从各种来源获得的大而复杂的数据集。许多最流行的数据处理技术都包含大数据技术，包括数据挖掘，机器学习，人工智能，数据融合，社交网络等[1]。许多人在各个领域使用大数据分析，并取得了巨大的成功[2]。例如，在商业领域，一些企业使用大数据更准确地了解消费者行为，从而优化产品价格，改善运营效率和降低人员成本[3]。在社交网络领域[3]，通过即时消息，在线社交网络，微博和共享空间的大数据分析，Facebook，Twitter和Linkedin等一些公司可以了解用户当前的行为，社交关系和社交行为规则，以及然后推广一些产品。在医疗保健领域，通过处理和查询医疗保健数据，医生可以分析病原体特征，评估患者的体质，从而制定更加人性化的治疗方案和建议，降低患者的发病率[4]。在智能电网领域，通过对智能电网数据的分析，电网运营商可以知道电力负荷和电力频率的哪些部分过高，甚至可以诊断哪些线路处于故障状态。这些数据分析的结果可以促进电网的升级，改造和维护工作[5]。随着大数据分析在如此多的领域中的成功应用，智能交通系统也开始以极大的兴趣关注大数据。

自20世纪70年代初以来，智能交通系统（ITS）得到了发展。这是交通系统的未来发展方向。ITS将先进的技术，包括电子传感器技术，数据传输技术和智能控制技术融入运输系统[6]。ITS的目的是为运输系统中的驾驶员和乘客提供更好的服务[7] - [9]。

在ITS中，数据可以从各种来源获得，例如智能卡，GPS，传感器，视频检测器，社交媒体等。对看似无组织的数据进行准确有效的数据分析可以为ITS提供更好的服务[10]，[11]。随着ITS的发展，ITS中生成的数据量正在从万亿字节级发展到Petabyte级。鉴于数据量如此之大，传统的数据处理系统效率低下，无法满足数据分析的要求。这是因为他们没有预见到数据量和复杂性的快速增长。

大数据分析为ITS提供了一种新的技术方法。ITS可以从以下几个方面受益于大数据分析。

1.大数据分析可以处理ITS中生成的大量不同且复杂的数据。大数据分析解决了三个问题：数据存储，数据分析和数据管理。Apache Hadoop和Spark等大数据平台能够处理大量数据，并且已广泛应用于学术界和工业界[12]，[13]。

2.大数据分析可以提高ITS运营效率。ITS中的许多子系统需要处理大量数据以提供信息或提供管理流量的决策。通过快速数据收集和当前和历史海量交通数据的分析，交通管理部门可以实时预测交通流量。公共交通大数据分析可以帮助管理部门了解交通网络中的乘客旅程模式，这可以用于更好的公共交通服务规划。运输APP开发人员的大数据分析可以帮助用户以最合适的路线以最短的时间到达目的地。

3.大数据分析可以提高ITS安全水平。使用先进的传感器和检测技术，可以获得大量的实时运输信息。通过大数据分析，我们可以有效地预测交通事故的发生。当事故发生或需要紧急救援时，基于大数据分析的系统中的实时响应能力可以极大地提高应急救援能力。大数据分析还可以提供识别资产问题的新机会，例如路面退化，压载老化等。它可以帮助在适当的时间做出维护决策，并防止车辆或基础设施处于故障状态。

虽然大数据分析在ITS中的应用具有很好的视野，但仍有许多关键的研究问题和重大挑战需要解决。据我们所知，从数据源和收集方法，数据分析方法和平台到ITS中的大数据分析应用程序的大数据分析的系统总结以前还没有完成。在本次调查中，我们首先讨论了ITS中大数据的来源以及我们如何收集生成的大数据。讨论了在ITS中进行大数据分析的框架。我们还总结了ITS中的数据分析方法和平台。还介绍了ITS中大数据分析应用的一些案例研究。

其余论文的结构如下。第二部分讨论了在ITS中进行大数据分析的架构。第三节总结了数据来源和收集方法。大数据分析方法在第四节中讨论。第五节详细介绍了ITS大数据分析应用案例研究。我们在第六节介绍了大数据分析平台。第七节讨论了在ITS中使用大数据分析的一些开放性挑战。最后，我们在第八节总结了这篇论文。

# 二、在其中进行大数据分析的体系结构

## A. ITS中的大数据特征

智能交通系统将先进技术融入交通系统[6]，其中包括电子传感器技术，数据传输技术和智能控制技术。ITS的目的是为运输系统中的驾驶员和乘客提供更好的服务[7]。根据[7]，ITS包括六个基本组成部分：先进的运输管理系统，先进的旅行者信息系统，先进的车辆控制系统，商务车辆管理，先进的公共交通系统和先进的城市交通系统。文献综述[7] - [9]表明这些组件中的大多数都是特定于车辆和公路运输的。因此，我们在本调查报告中关注ITS的公路运输。

智能交通系统（ITS）收集的数据越来越复杂，并且具有大数据功能。包括Gartner IBM和微软在内的大公司提出，大数据可以通过三个V来描述，即数量volume，种类variety和速度velocity [14]，[15]。

数量是指各种来源产生的数据量，并且仍在扩大。随着交通量和探测器数量的增加，运输中的数据量显着增加。此外，在使用跟踪转发器时，旅行者，货物和车辆会产生更多数据。基础设施，环境和气象监测产生的数据也正在成为运输数据的关键部分。

多样性主要集中在由探测器，传感器甚至社交媒体产生的各种数据上。各种与运输有关的数据显着增加。例如，现代车辆可以实时报告内部系统遥测以及所有机组成员和乘客的信息。

由于通信技术的改进，处理能力的提高以及监测和处理的速度，运输中的数据速度有所提高。例如，现在立即报告使用智能卡或标签的票务和收费交易，而基于纸张的票务需要人工处理以从交易中获取有用的数据。

## B.在ITS中进行大数据分析的体系结构

在ITS中进行大数据分析的体系结构如图1所示。它可以分为三层，即数据收集层，数据分析层和应用层。

•数据收集层：数据收集层是架构的基础，因为它为上层提供了必要的数据。数据来自各种来源，如感应环路探测器，微波雷达，视频监控，遥感，射频识别数据和GPS等。有关大数据收集的详细信息将在下一节中介绍。

•数据分析层：数据分析层是架构的核心层。该层主要用于从数据收集层接收数据，然后应用各种大数据分析方法和相应的平台来完成数据存储，管理，挖掘，分析和共享。有关大数据分析方法和平台的详细信息将在下一节中介绍。

•应用层：应用层是此架构中的最顶层。它在不同的运输环境中应用来自数据分析层的数据处理结果，例如，交通流量预测，交通引导，信号控制和紧急救援等。

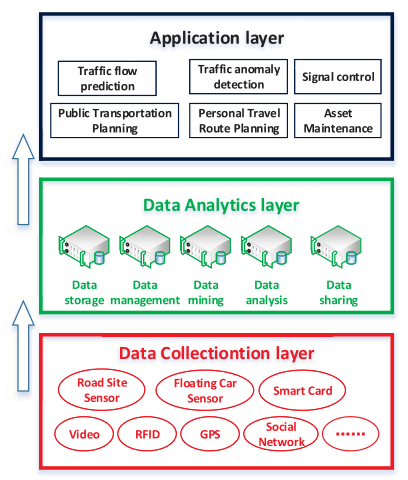


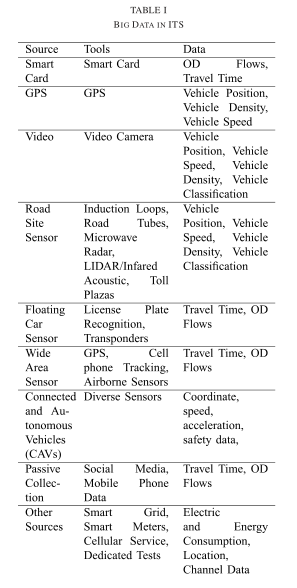
图 1 在ITS中进行大数据分析的架构

数据采集层使用先进的数据采集技术监控人员，车辆，道路和环境。包括结构化数据，半结构化和混合数据的原始交通数据通过有线或无线通信传输到数据分析层。在数据分析层接收到原始流量数据后，它首先对数据进行分类，删除重复数据，清理数据并以分布式方式分发有用且准确的数据。然后利用数学和工程理论提取隐藏信息，主要包括描述性分析和预测分析。利用分析结果，应用层可以预测未来交通流量和乘客流量的趋势，分析交通事故多发地点，调整信号分布，实施交通控制，为城市管理部门提供决策支持。

# 三、其中的大数据收集

人们无意识地参与了ITS中大数据的收集，传输和应用。ITS中的技术发展导致从车辆和人员流动中创建和收集的数据的复杂性，多样性和数量的增加。根据ITS中的不同来源，ITS中的大数据可以主要分为以下类型，收集的数据如表I所示。

表 1 ITS中的大数据



## A.智能卡的大数据

自动票价收集（AFC）系统已广泛应用于城市轨道交通系统，这使得智能卡数据成为调查乘客运动模式的主要数据来源[16] - [18]。在AFC系统中，乘客在乘坐公共汽车或火车时需要使用智能卡。电子阅读器将在触摸智能卡时捕获乘客详细信息，如登机时间，OD信息等。AFC系统中的智能卡每天都会在大城市中生成大量的数据记录。例如，伦敦交通局（伦敦交通局）每天从伦敦地铁站的800万次旅行中收集智能卡数据。

已经做了大量工作来使用智能卡数据来研究公共交通乘客出行行为的空间和时间模式[19] - [22]。由于其具有提供旅行行为综合时空信息的潜力[17]，[21]，智能卡数据正成为公共交通服务规划和管理的重要组成部分。

## B.来自GPS的大数据

GPS是最受欢迎的位置跟踪工具。通过GPS进行位置跟踪，可以更有效，更安全地收集交通数据。结合地理信息系统（GIS）或其他地图显示技术，GPS提供了一种有前途的数据收集工具，收集的数据可用于解决许多交通问题，如旅行模式检测[23]，[24]，旅行延误测量[25]和交通监测[26]。

## C.视频中的大数据

摄像机广泛部署在ITS中。如先进的交通管理系统（ATMS）所示，与传统的传感器相比，视频图像检测系统（VIDS）是车辆识别和交通流检测等任务的理想选择。 VIDS的一个优点是成本低[27]。已成功部署使用海量视频数据的高速公路成像传感器进行事件检测，并在某些情况下显示出高精度[28]。除了一般的交通管理[29]，收集更准确的车辆视频数据的运输工程师和规划人员可以改进图像处理系统，以便更好地制定关于车辆排放模型的一般运输需求。

## D.来自传感器的大数据

ITS中安装的传感器设备用于收集车辆速度，车辆密度，交通流量和行程时间等数据。传统的道路传感器（例如，红外和微波探测器）已经不断发展以获取，计算和传输交通数据[30]。如[30]所示，传感器的数据收集可分为三个来源：路边数据，浮动汽车数据和广域数据[31]。

路边数据主要是指位于路边的传感器收集的数据。传统的路边传感器，例如感应磁环，气动路管，压电回路阵列和微波雷达已经使用了很多年。新一代路边传感器，如超声波和声学传感器系统，磁力计车辆探测器，红外系统，光探测和测距（LIDAR），以及视频图像处理和探测系统，随着最近的先进技术发展逐渐出现。

浮动车数据（FCD）主要是指ITS中不同位置的车辆移动数据，其中定制的探测器嵌入车辆中[32]。一些板载传感器为旅行路线选择和估计提供了有效和有效的信息。随着车辆传感器技术的发展，流行的FCD传感器技术包括：自动车辆识别（AVI），车牌识别（LPR）以及诸如探测车辆和电子收费标签的转发器。

广域数据是指通过各种传感器跟踪技术（如摄影测量处理，录音，视频处理和空基雷达）收集的广域流量数据。

## E.来自CAV和VANET的大数据

联网和自动驾驶车辆（CAV）是ITS领域的新技术，它结合了车辆设计的根本变化以及它们与道路基础设施的相互作用。联网和自动驾驶汽车采用了一系列不同的技术，促进了人员和货物的安全，有效运输。启用CAV的交通系统已显示出缓解拥堵，减少行程延误和提高安全绩效的巨大潜力[33]，[34]。CAV可以生成大量与环境相关的实时运输数据，例如坐标，速度，加速度，安全数据[33]。使用软件定义网络等最新网络技术，可以更有效地获取数据[35]这些数据可用于创建可操作的信息，以支持和促进绿色交通选择，并应用于实时自适应信号控制[36]，[37]。

车载Ad Hoc网络（VANET）是一种移动ad hoc网络，它使用车辆和基础设施元素作为节点来增加覆盖区域和通信能力。作为ITS的重要组成部分，VANET会产生大量数据[38]。数据准备和实时结果是大规模分析的挑战性任务。使用大数据分析，我们可以解决大多数与VANET相关的数据挑战[39]，例如数据填充[40]，拥塞和事故警报[41]以及交通流量预测[42]。

## F.被动收集的大数据

与运输研究中积极收集的数据相比，移动技术的快速发展使得能够收集大量的被动数据。被动数据是指未通过主动收集收集的数据。它的生成目的不是有意的，但可以用于研究[43]，[44]。陈等人[45]和Zeyu等人[46]建议结合被动大数据，如手机数据，互联网访问数据和活动数据，以研究人员流动性，旅行行为和交通规划。在[47]中，诸如当前时间，手机ID，用户身份的上下文信息用于预测移动用户的停留时间。

社交媒体数据是最受欢迎的被动数据，它指的是人们互相交互以创建，共享和交换信息和想法的应用程序或网站。最近，Linkedin，Facebook和Twitter等社交媒体网络迅速发展起来。它们已成为运输专业人员的相关利益，因为它们实时地在供应商和消费者之间提供信息流[48]。虽然通过社交媒体网络收集的数据通常是非结构化的并且需要复杂的处理，但是当在不同类型的交通中表达态度时，它提供了重要的交通信息，并且在社交媒体中发现了对旅行中断的响应[49] - [52]。

## G.来自其他来源的大数据

有些数据来源无法分类为上述类别。例如，实时基础设施状态被认为是重要的数据来源[53]。最著名的例子是智能电网[54]，它将使我们能够收集电动汽车的日常用电信息和城市轨道交通系统中的列车牵引力。另一个重要的数据来源是ITS中专用测试的数据。例如，在我们之前的工作中，我们在城市轨道交通基于通信的列车控制（CBTC）系统[55]，[56]的真实列车地面通信系统中进行现场测试。从现场测试中获得大量的信道增益数据。处理数据以模拟信道状态的随机特性，并且该模型用于优化CBTC系统性能。

# 四、大数据分析方法

机器学习是大数据生态系统中最流行的建模和分析理论，可以轻松地从大量数据中获取模式和模型。在ITS领域，机器学习理论也被广泛用于进行数据分析。根据可用于学习的数据集的完整性，机器学习模型可以分为监督，无监督和强化学习算法。随着近年来人工智能的快速发展，最近ITS也采用了强大的深度学习模型。

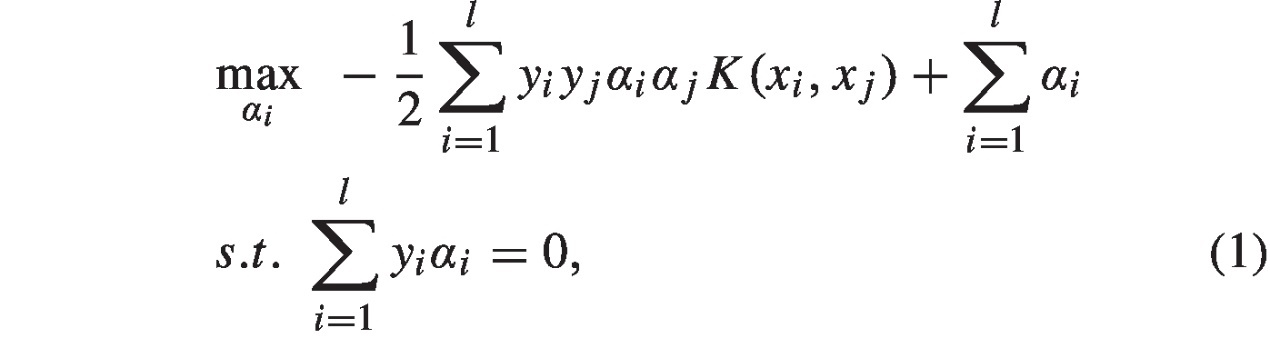
## A.监督学习

标记的训练数据用于监督学习算法[57]。模型使用输入数据和目标输出（标签）来学习它们之间的功能或映射。结合学习模型和输入数据，可以预测看不见的输出。在所有监督学习模型中，线性回归，决策树，神经网络和支持向量机是ITS中最常用的。

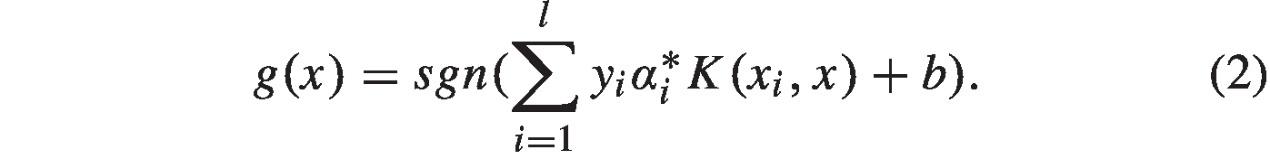
回归函数用于解释一个因变量与一个或多个自变量之间的关系。线性回归是最常用的监督学习[58]。线性回归非常简单，强大，易于理解，易于编码。尽管简单，但线性回归在各种ITS情景中尤其成功，例如交通流量预测[59]，交通速度估算[60]和交通出行路线评估[61]。决策树是一种决策支持工具，它使用树状图来模拟决策及其可能的后果[62]。由于其可移植性，鲁棒性和透明性，决策树广泛用于各种ITS场景，例如交通事故检测[63]，事故严重性分析[64]和旅行模式选择[65]。

人工神经网络（ANN）是分类和回归的灵活有力的监督学习的一个流行的例子[57]。通过处理节点和训练数据的足够隐藏层，ANN可以学习输入和目标数据之间的任何非线性关系。作为一种数据建模工具，它也被ITS采用，如交通流预测[66]，行程时间预测[67]，交通事故检测[68]和剩余停车位预测[69]。支持向量机（SVM）是另一种流行的监督学习算法，它使用标记数据进行回归和分类。在ITS中的所有大数据分析模型工具中，SVM引起了研究领域的极大兴趣。它已成功用于旅行时间预测[70]，公交车到达时间预测[71]和交通事故检测[72]。

在[72]中引入了在ITS中使用监督学习的典型示例，其中SVM用于预测交通事故。给定训练子集，其中是训练样本的输入，其包括交通流量参数的值，例如音量，速度，占用率等，以及是的类标签。利用核函数，根据SVM分类器理论，支持向量可以得到，



然后，我们得到决策函数来计算样本的标签，



如果x是一个入射样本，。否则，我们有。

## B.无监督学习

无监督学习通常也称为聚类，重点是从未标记的多维数据中学习自然群[57]。K-means是最受欢迎的无监督学习工具，它已被广泛应用于公路交通规划[73]和旅行时间预测[74]。通过一组历史数据，[74]的作者给出了使用无监督学习来预测旅行时间的经典例子。程序如下，

1.计算行程时间频率ε。它表示旅行时间出现的时间。

2.定义一个元组包含不同的特征，其中是行程时间，是行程时间频率，是行进速度。

3.根据旅行时间频率找出数据中的最大值。选择元组作为簇1的质心，其中是最大行程时间频率，是与相关的相应最大行程时间，是与之相关的行进速度。

4.通过计算它们的距离，比较群集1的每个元组与质心。选择元组最大距离。

5.构建两个簇，其中簇1的质心是元组和簇2的质心是元组。

6.通过将所有元组分配给最近的集群质心来定义所有元组的集群成员资格。

7.使用算术平均值重新估计群集中心。

8.重复步骤6和7。

9.在完成簇的准备之后，针对每个簇分别计算期望的预测时间，，其中是从第k个聚类获得的旅行时间，N是相关聚类中元组的总数，是第j个元组的旅行时间频率，是第j个元组的旅行时间。

10.最终预测的近似旅行时间是通过计算和的算术平均值得到的。

## C.强化学习

与有监督和无监督学习不同，如图2所示，强化学习的目的是通过探索最小化长期成本。

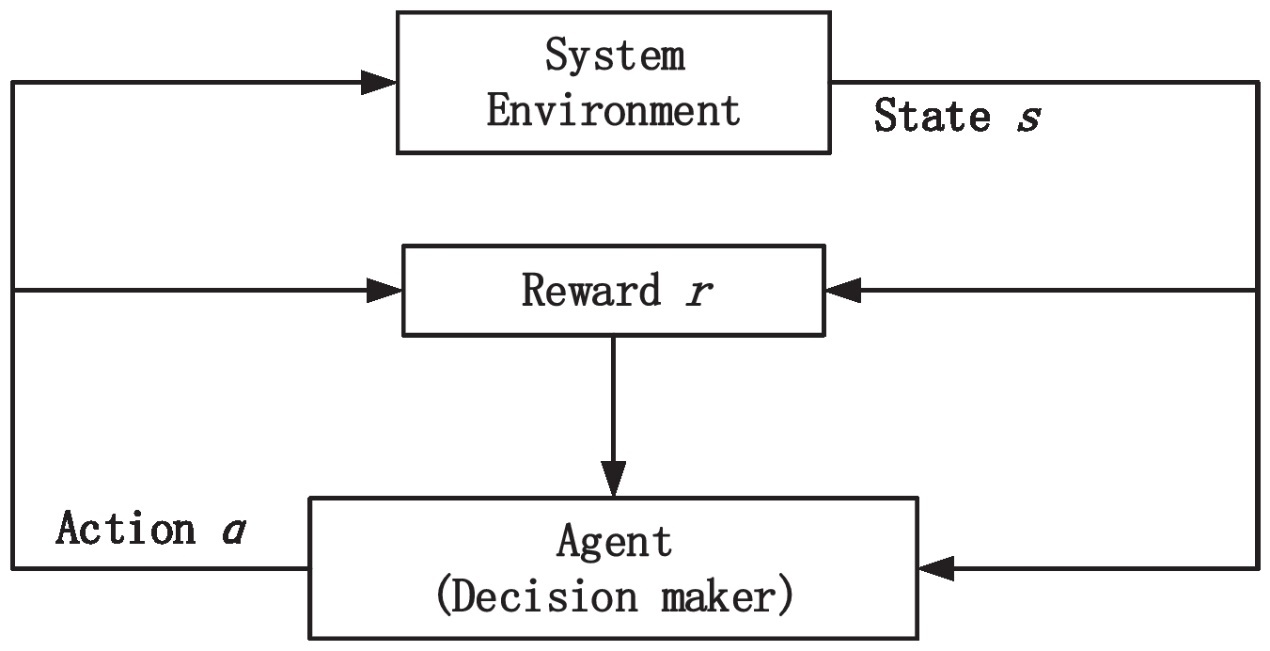


图 2 强化学习

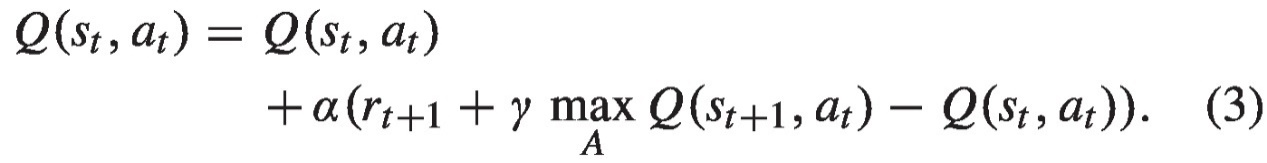
通过与实验数据相互作用来学习最优政策[57]。强化学习与控制和优化理论高度相关，并且已经证明它在ITS中的交通信号控制中是非常可行的[75] - [78]。

在ITS中使用强化学习需要在强化学习的语言中制定ITS控制和优化问题，具体地说，定义状态空间S，动作空间A和奖励R.在[76]中制定了在ITS交通信号控制中使用强化学习的一个典型例子。在具有n个通道的交叉点处的交通状态被正式定义为离散交通状态编码（DTSE）。对于接近交叉路口的每个车道，DTSE将从停车线开始的车道段的长度l离散为长度为c的单元。选择c将改变系统的行为。 DTSE由三个向量组成。第一个向量B表示在细胞中是否存在载体。第二矢量R表示车辆的速度，第三矢量P是当前的交通信号相位（即，所选择的最近的动作）。因此，系统状态可以定义为

在代理观察到环境状态后，它必须从所有可用操作集中选择一个操作。可能的行动是南北绿（a1），东西绿（a2）），南北推进左绿（a3）），东西推进左绿（a4））。所有可能的动作A的集合被定义为A={a1，a2，a3，a4}。在时间t，代理选择动作a（t），其中。

在代理观察到环境st的状态之后，它执行动作a（t），并接收奖励。奖励rt + 1是从特定状态执行选定动作的结果。在此公式中，奖励被定义为行动之间累积车辆延迟的变化。

该公式中使用的强化学习算法是Q-Learning [57]，用于制定最优的动作选择策略。通过使用卷积神经网络来近似动作 - 值函数来实现最优策略。动作值函数Q（st，at）将状态映射到动作实用程序（即，来自给定状态的每个动作的值是多少）。Q学习的基础是价值迭代更新定义为，



学习率α控制新行动 - 价值估计与旧估计加权的程度，折扣因子γ确定如何立即对未来奖励加权。在充分学习了动作值功能之后，可以通过选择具有最高值的动作来确定最优策略。

## D.深度学习

深度学习模型比传统的人工神经网络利用更多的系统特征和复杂的架构，并且可以实现比传统机器学习模型更好的性能。它们已被广泛应用于ITSs。例如，利用深度限制玻尔兹曼机器和递归神经网络架构来模拟和预测基于出租车的GPS数据的交通拥堵演变[79]。使用深度神经网络，在大数据的转向架上进行故障诊断在[80]中进行。Chen [81]利用从ImageNet数据集中学习的卷积神经网络（CNN）的丰富特征来执行车辆检测任务。段等人 [82]使用堆叠式自动编码器进行交通数据插补。在交通流域，深度学习模型已成为预测交通流密度的流行工具[83] - [86]。文献[85]给出了一种典型的基于深度学习的方法来进行交通流量预测。堆叠自动编码器（SAE）用于学习通用流量传输功能。考虑具有K层的SAE，第一层被训练为自动编码器，训练集作为输入。在获得第一隐藏层之后，第j个隐藏层的输出被用作第（k + 1）个隐藏层的输入。以这种方式，可以分层堆叠多个自动编码器。为了使用SAE网络进行交通流量预测，在网络顶部添加了逻辑回归层，用于监督交通流量预测。整个深层架构模型如图3所示。

从所有高速公路收集的数据用作输入。考虑到交通流量的时间关系，交通流量数据在先前的时间间隔，即。用于预测时间间隔t的交通流量。所提出的模型固有地解释了交通流的时空相关性。

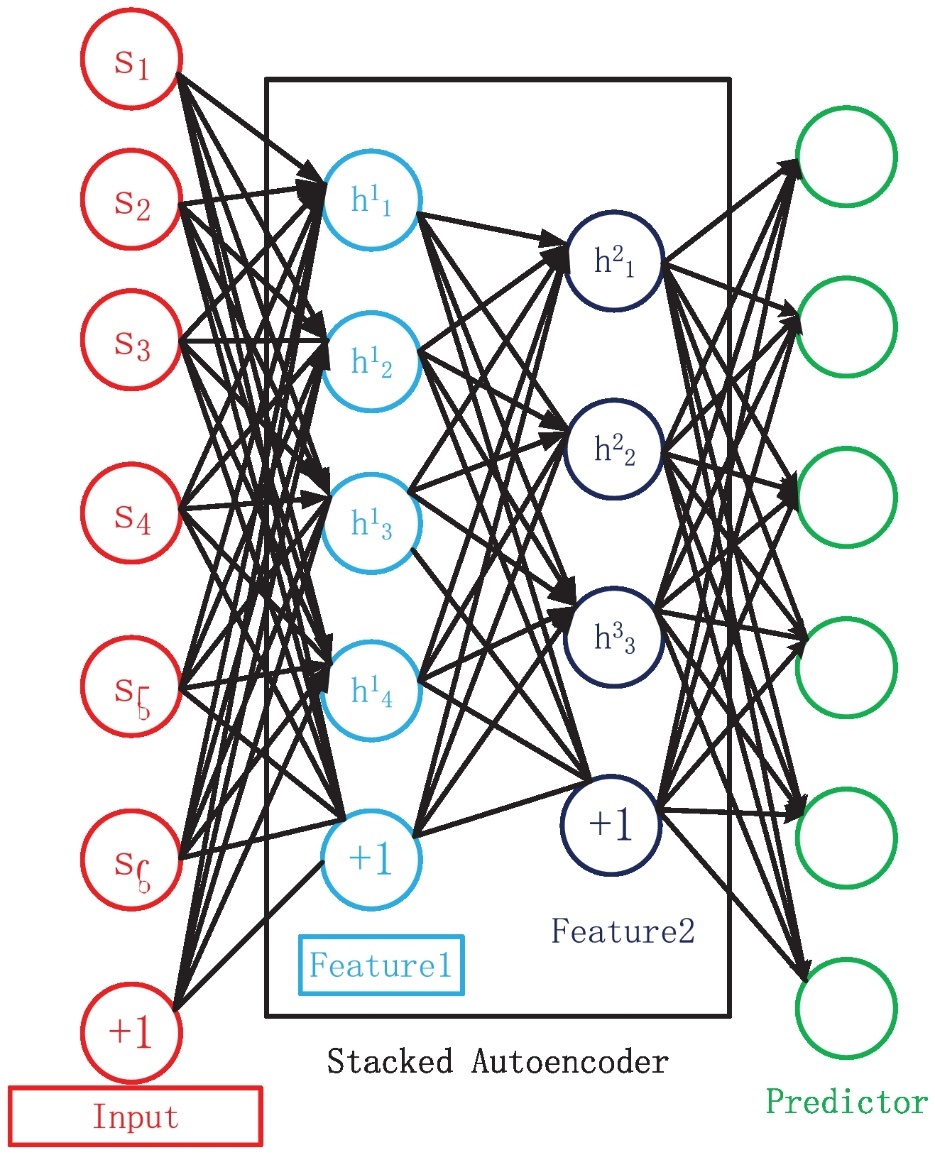


图 3 交通流预测的深度架构模型

## E.基于本体的方法

本体论是对特定话语领域中存在的实体的类型，属性和相互关系的正式命名和定义。基于本体的方法可以准确地描述数据语义并推断隐式数据语义关系。与自下而上的传统数据提取相比，本体数据集成具有自上而下的特征，并使用本体建模来共享数据的语义视图并映射来自不同数据源的异构数据，以最小化甚至消除对共享数据的模糊理解。基于Ontology的方法已在ITS中得到广泛应用。例如，Zhai等人 [87]提出了一种基于模糊本体框架的ITS信息检索系统。该框架包括三个部分：概念，概念属性和属性值，并侧重于有关交通事故的信息。Fernandez和Ito [88]提出了一种基于本体的智能交通系统驾驶员行为模型。驾驶员行为本体具有与驾驶员特征，感知和认知状态相关的知识以执行不同的驾驶任务。它可用于预测交通事故并优化道路拥堵。Fernandez和Ito [89]建议使用本体来管理智能交通系统中的传感器信息，并将传感器数据转换为语义数据。系统执行自动交通灯设置可以使用数据来预测和避免交通事故。格雷戈尔等人 [90]提出了一种在ITS领域创建本体的系统方法。该本体将作为语义服务的语义信息的基础，该语义服务允许将新设备连接到城市网络。赵等人 [91]介绍了一个基于本体的知识库，其中包含地图和交通法规。通过访问知识库，智能车辆可以了解超速情况，并在交叉路口做出符合交通法规的决策。陈等人 [92]描述了一种基于本体的协同ITS（C-ITS）安全管理方法，主要用于汽车领域。它为本体驱动的ITS开发及其正式的信息模型提供支持。Yang和Wang [93]利用本体的语义完备性来构建城市交通本体模型，解决了交通信息集成语义融合中的本体合并和等价验证等问题。该模型可以增加语义融合的功能，减少城市交通信息的数据整合量，提高交通信息查询的效率和完整性。

# 五、大数据应用

大数据为ITS的开发和应用提供技术支持。通过高效，准确，及时的公路和铁路运输系统数据采集，分析和处理，大数据应用可以为公众提供方便，高效的运输。为了识别问题，提高ITS效率，降低成本并获得有价值的见解，ITS中的大数据应用可分为以下六类。

## A.道路交通事故分析

有证据表明，全世界每年约有120万人死于交通事故，造成5000万人受伤[94]。准确的交通事故数据分析结果可以为交通部门提供重要信息，以制定政策以防止事故发生。

许多研究都侧重于在交通事故分析中使用大数据分析。使用测量的交通流量数据，Golob和Recker [95]利用多元统计模型研究了天气，光照条件，交通流量和城市高速公路事故之间的关系。在[10]中，在实时碰撞预测模型中采用贝叶斯推理和随机森林来降低碰撞风险。熊等人[96]引入分类和回归树（CART），逻辑回归和多元自适应回归样条（MARS）来执行机动车事故伤害数据的分析操作。Lee和Mannering [97]提出了一种方法，该方法使用零入射计数模型和嵌套logit模型来分析华盛顿州96.6km高速公路上的径流事故频率和严重程度。结果表明，可以采取一些措施来减少径流事故频率。Karlaftis和Golias [98]采用严格的非参数统计方法，即基于分层树的回归（HTBR）来分析地形和交通特征对农村道路事故率的影响。该方法还可用于预测高速公路的事故率。张，等。通过使用负二项回归模型和分类和回归树模型分析公路几何变量与交通事故之间的关系。这些参数来自台湾国家高速公路1号的2001 - 2002年事故数据[99]。Bédard等 [100]通过使用多变量逻辑回归算法确定驾驶员，碰撞和车辆特征对驾驶员死亡风险的相应影响，结果表明增加安全带使用，降低车速，减少驾驶员的数量和严重程度侧面撞击可以防止交通事故。

## B.道路交通流量预测

及时准确的交通信息对于运输管理至关重要。ITS中的大数据分析在交通流量预测方面具有优势[101] - [103]。根据[9]，使用大数据分析的经典道路交通流预测模型如图4所示。首先对原始ITS数据进行预处理以获得有效数据集。使用选定的数据挖掘或分析方法，使用预处理数据建立交通流模型。交通流模型为交通管理部门提供决策支持，并从实际交通流中获得反馈，以校准模型。

许多学者使用大数据分析研究了交通流量预测。Lv等人 [85]基于图4提出深度学习。

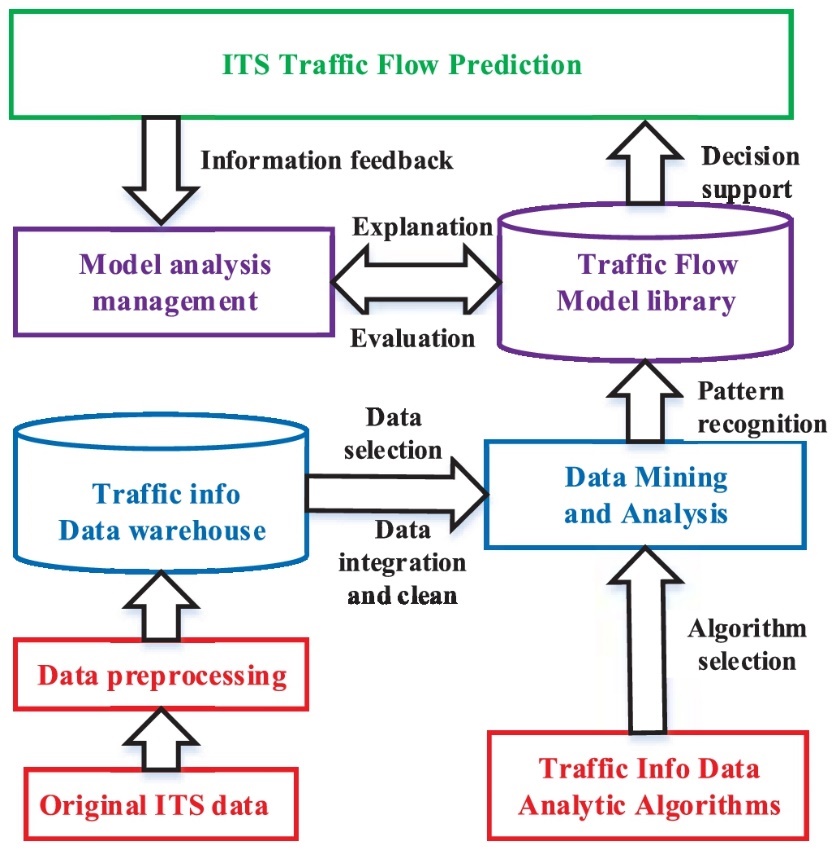


图 4 典型的交通流预测模型

利用贪婪分层无监督学习算法的交通流预测方法。堆叠式自动编码器（SAE）模型用于学习通用流量传输功能。结果表明，基于深度学习的模型在交通流预测中具有优越的性能。刘等人 [104]分析多维参数，交通流预测模型是基于SVM从不同维度开发的。董等人 [105]提出了一种预选空间时间模型，用于估算数据检测器较少的位置的交通流量。Canaud等人 [106]提出了一种基于概率假设密度滤波的模型，用于实时交通流量预测。潘等人 [107]提出了一种改进的随机细胞传递模型，以支持短期交通流量预测。Antoniou等。 [108]提出了一种基于数据驱动计算方法的局部流量状态估计和预测方法。使用看似无关的时间序列方程（SUTSE），Ghosh等。 [109]提出了一种新的多变量结构时间序列（MST）模型来预测交通流量。随着时间的推移，SUTSE模型可以分别跟踪每个交通流量及其组成部分的变化，结果表明它具有更高的预测精度。徐等人 [110]提出了一种新颖的在线算法，它是一种上下文感知自适应交通预测算法。该算法可以从当前的交通状况中学习，并使用历史交通数据来预测未来的交通流量。实验表明，该算法比目前的解决方案做得更好。Lu等人 [111]建立一个交通流状态聚类模型，采用模糊退火遗传算法，采用模糊c-均值（SAGA-FCM）。该模型基于交通速度数据和占用数据，全面考虑交通流量大数据的时间，空间和历史相关性。

随着近来AI技术的快速发展，深度学习方法已被广泛应用于预测交通流。黄等人 [84]将深厚的信念网络引入交通系统。Ma等人 [79]将深度限制玻尔兹曼机（RBM）与RNN相结合，形成了RBM-RNN模型，该模型继承了RBM和RNN的优点。他们还[86]使用LSTM预测交通，并证明LSTM与传统神经网络相比，通过使用北京公路网收集的环路探测器数据，在稳定性和准确性方面实现了更好的性能。Lv等人 [85]提出了一种新的基于深度学习的交通预测模型，该模型考虑了时空关系，并采用堆栈自动编码器（SAE）来提取交通特征。

## C.公共交通服务规划

公共交通大数据分析有助于了解交通网络中的交通乘客旅程模式。乘客行程模式可用于向运输运营商通报有关服务计划的决策。

对于交通测量数据的异构来源，Lu等人 [112]提出了一种基于路径流的非线性优化模型，用于估计不需要显式动态链接信息的动态OD需求。使用数百万匿名用户的三角手机记录，[113]的作者提出了一种预测平均每日OD旅行的方法。所提出的模型的适用性通过来自地方和国家调查的旅行的空间和时间分布来验证。Gordon [114]使用来自伦敦地铁和iBus车辆定位系统的完整日常智能卡数据，得出每位乘客的登机和下车时间，并且转乘信息来自属于不同公共交通方式的旅客出行。完整的旅程矩阵是根据数据建立的，并由传统的O-D矩阵验证。该方法足以每天执行，并为运输运营商提供其服务的旅行行为。这些工作中的大数据分析结果可以帮助新兴的智能交通管理应用程序生成主动，协调的交通信息提供。Tao [115]研究了五次典型日历事件中快速公交（BRT）旅行对非BRT旅行的时空动态。首先对智能卡数据进行预处理，以分别为BRT和非BRT行程构建OD流量矩阵和总线跳闸路径。服务管理部门可以识别基于证据的BRT政策的重要含义。在[116]中，自动收费系统（AFC）系统的运营大数据用于土耳其伊斯坦布尔的交通规划管理。麻省理工学院的工作[117]显示了伦敦AFC数据在铁路运输规划和运营中的潜在价值。在他们的工作中开发的应用程序为铁路运输运营商和规划人员提供了一种易于更新的管理工具，可以近乎实时地评估几个方面的铁路服务。Toole等。 [118]使用来自开源数据存储库的移动电话数据来实现旅行需求模型。可以从模型中的呼叫数据记录（CDR）数据中提取可路由的道路网络，经验证的OD矩阵和行程表。他们的工作是帮助运输经营者执行公共交通规划的通用指南。

## D.个人旅行路线规划

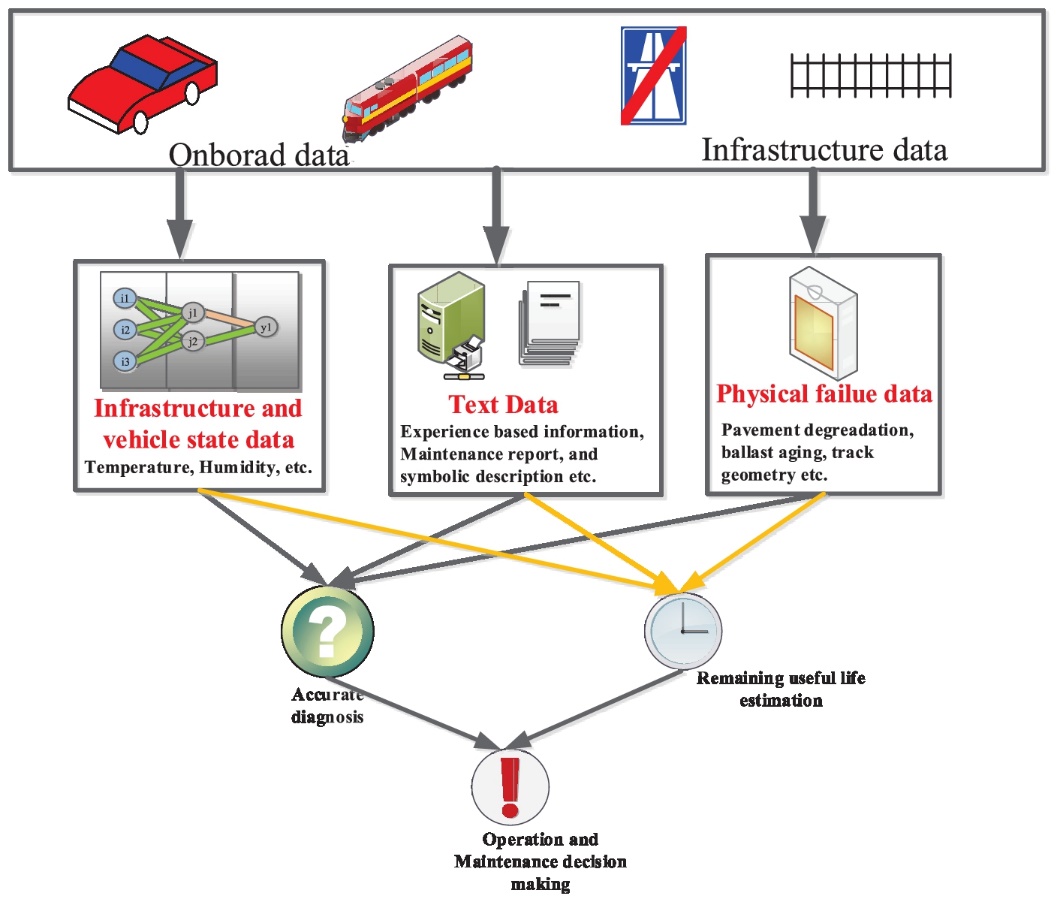
运输应用程序始于良好的愿景。报告建议只告诉乘客下一班车的到达时间可以使他们更满意公交服务[119]。根据智能手机和车辆GPS数据的数据，一些运输APP为乘客提供实时交通信息[120]，其他提供最合适的行车路线，最短的旅行时间[121]。通过智能手机将公共交通数据与用户信息相结合，交通APP甚至可以为乘客提供实时的公共交通旅程规划[122]。完全集成的应用程序甚至可以让人们计划从火车到公共汽车和私家车或自行车的旅程[123]。

这些运输APP中的大数据分析通过减少旅行时间，交通拥堵，污染和温室气体排放产生了巨大的经济效益。例如，开放伦敦交通局（TfL）数据的价值每年为15-58百万英镑，并且私人公司开发了200多个旅行应用程序[124]。

## E.铁路运输管理和控制

铁路运输系统已经通过先进的IT技术进行了改造。它们是大数据分析的主要优势。这是因为铁路运输系统通常是封闭系统，其执行大量数据的复杂处理，例如实时列车速度和位置，某个站的列车出发和到达时间以及乘客OD信息。大数据分析可以使铁路运输运营商更好地控制列车，提高铁路运输系统的运营效率。在工业领域，大数据分析开始在铁路运输系统中发挥重要作用。作为典型的公共铁路运输系统，湾区快速交通（BART）保持对其系统所有阶段的监督，包括列车运营，客运服务，供电和路旁设施。大数据分析是所有这些功能中的关键要素。Schultz在[125]中指出，BART运营分析的关键作用是确保进度可靠性。利用马尔可夫链模型，开发了伦敦的多式联运网络，为运输效率提供了更好的信息集群[126]。在[127]中，大数据分析应用于荷兰乌得勒支，以预测交通并通过移动电话，智能卡和计算机的数据改善运营。

在学术界，在铁路运输管理和控制中使用大数据分析方面已经做了大量工作。利用上海轨道交通1号线的客运OD信息，Jiang等。 [128]评估列车时刻表的效率。这种方法在真实的铁路运输系统中得到验证，该系统涉及超过100万人次的旅客和600列火车。尹等人 [129]提出了一种智能列车运行（STO）方法，其结合了自动列车运行（ATO）和手动驾驶的优点。将专家知识与数据挖掘算法融合，应用于STO方法。结果表明，在能耗和乘坐舒适性方面，该方法优于ATO，在准时性和停车精度方面，也优于手动驾驶。陈等人 [130]通过使用列车速度数据，提出了关于列车停止误差与列车控制参数之间关系的两个简化模型，如图5所示。使用大数据分析进行资产维护的典型框架。



停止前的时间数据和距离数据，并引入一种在线学习算法 - 多项式adaline算法，以提高停车精度。结果表明，所提出的简化模型和在线学习算法可有效地减少停车误差并纠正列车停车误差分布的偏差。周[131]应用两种典型的机器学习算法高斯过程和Boosting，通过利用停止前的一些初始速度数据和距离数据来提高列车停止精度，结果表明高斯过程回归算法获得了最佳性能。侯等人 [132]提出了三种列车停止控制算法，其选择初始制动位置数据，制动力数据及其组合数据作为控制输入。基于终端迭代学习控制（TILC），这些算法使用先前制动过程中的停止位置误差来提高列车停止精度。陈等人 [133]使用新的机器学习技术，并提出新颖的在线学习控制算法，以实现列车自动停止控制。该算法包括启发式在线学习算法（HOA），基于梯度下降的在线学习算法（GOA）和基于RL的在线学习算法（RLA）。所需参数来自轨道侧应答器。结果表明，该方法可以在常规干扰下将停止误差限制在±0.30m范围内。

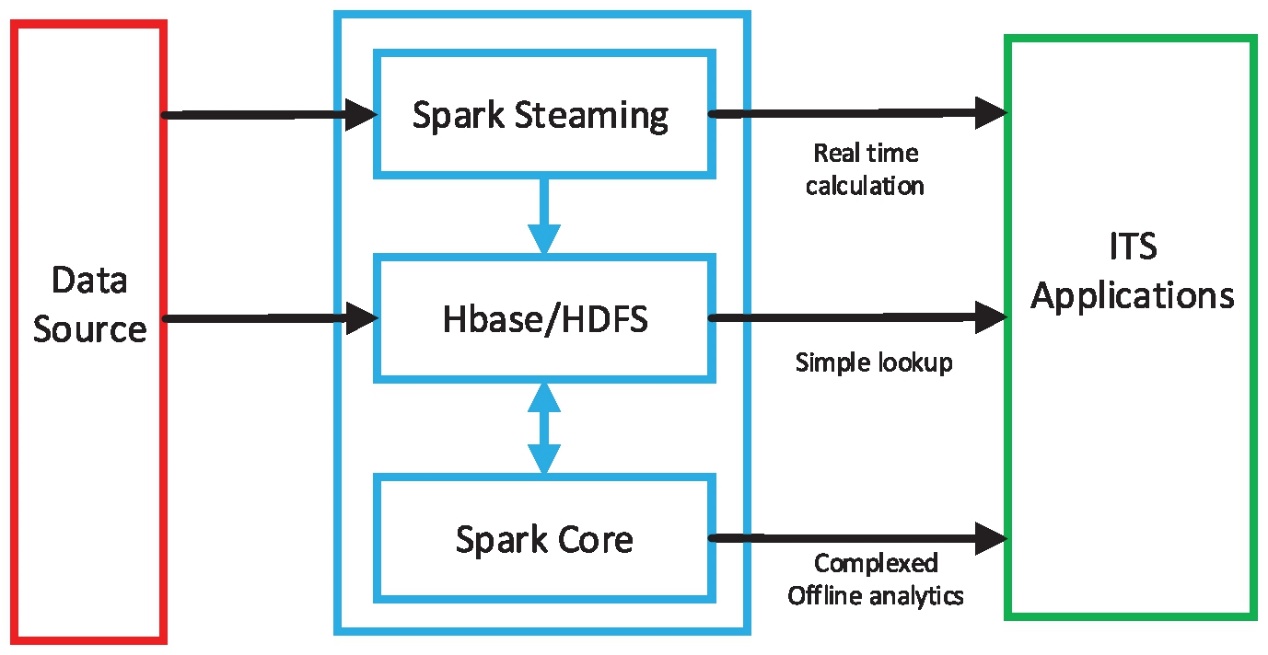
## F.资产维护

在ITS中，有大量资产依赖于大量数据来运营和维护。适当的资产维护方法对于保护ITS资本和降低维护成本非常重要。大数据分析可以帮助更快，更准确地识别问题，并最大限度地降低维护成本。使用大数据分析进行资产维护决策[134]的典型框架如图5所示。从不同的传感器收集板载和基础设施数据。可以直接使用诸如路面退化，压载物老化，轨道几何形状等物理故障数据。可以处理诸如基于经验的信息和维护报告，符号描述等的文本数据以提取重要信息。可以使用数据驱动方法处理诸如温度，湿度等的基础设施和车辆状态数据，并获得状况指标。整合三种过程方法的结果，获得资产状况的准确诊断，确定资产的剩余使用寿命，可供最终用户做出维护或操作决策。基于大数据分析的维护的一个示例由荷兰铁路公司在Axle Box Acceleration（ABA）上进行。利用1TB的轨道退化数据，可以执行自学习和适应机制[135]。Thaduri等人 [134]引入混合建模方法，以提供准确的诊断资产条件，以确定资产的剩余使用寿命。所提出的方法提供了为最终用户提供维护决策的洞察力。基于语义数据模型，铁路资产监控系统在[136]中实施，并证明与传统方法相比，更能够实现数据集成，可扩展性和兼容性。Zarembski [137]利用多条检查车辆在330,000公里的铁路轨道上收集的数据，介绍了数据采集，存储和使用大数据分析计划铁路轨道维护的程序，以优化其资本基础设施并控制成本。使用来自智能手机和GPS坐标的数据，英国的Network Rail成功地将轨道缺陷位置从1英里提高到5米，这大大缩短了修复铁路轨道的时间[138]。利用大量的历史系统状态数据，结合列车类型数据，维护行动数据，检查计划数据和系统故障数据，Li等人 [139]探索了几种基于机器学习的分析方法，以自动学习规则并构建故障估计模型。模型可以使用实时数据来估计当前条件是否会导致系统故障。在[140]中提出了一种基于双层特征提取的文本挖掘方法，其中使用了在语义和语法级别提取的特征。该方法显着提高了所有故障类别的故障诊断精度。

# 六、其中的大数据平台

ITS中的大数据分析在先进的大数据平台的帮助下不断发展。大数据平台利用分布式文件系统和并行计算功能来实现快速数据处理。它能够理解大数据并支持大规模系统优化。

Apache Hadoop是最流行的开源软件框架，用于分布式处理和存储大量数据集。Hadoop是一个通用的大数据处理平台，可以执行各种数据处理或数据分析操作。分布式处理能力使Hadoop非常适合分析ITS中的数据，如智能卡数据，各种传感器，社交媒体，GPS数据等。Apache Spark是用于大量数据集处理的最新开源平台，特别适用于机器学习任务[141]。Spark采用与Hadoop相同的分布式存储技术，允许用户使用图6。在ITS中使用Apache Spark平台的典型框架。



将数据加载到集群内存并重复查询的程序。Spark非常适合机器学习方法。我们在上一小节中介绍的大数据分析方法是基于机器学习的，它们可以在Hadoop和Spark平台上进行定义。运行数据分析方法的大数据平台将在ITS中的大数据分析中发挥重要作用。

在ITS中使用Spark平台的典型框架如图6所示。来自不同来源的数据由HBase（Hadoop数据库）API收集，并发送到数据中心。 Spark Streaming实时处理数据。可以实现一些实时任务，例如车速检测，车辆识别，实时警告等。 HBase是一个分布式开源数据库。它将执行高级特征提取，并为海量数据集创建索引，从而提高数据检索的有效性和效率。Spark Core是火花系统的基础，它可以执行具有分布式计算功能的离线任务。可以在Spark Core引擎下进行交通管理和控制，事故分析等关键任务。

与通用大数据平台不同，在交通领域，已经提出了几个平台来处理运输数据。

Mian等[142]提出了一个具有多个引擎的平台，以支持各种类型的交通数据分析。Zareian等人 [143]提出了一个名为K-Feed的监控系统，用于对部署在云上的应用程序进行性能分析。斯特恩等人 [144]提出了一个数据引擎哥斯拉的概念架构，以执行实时交通数据处理并支持对运输数据的分析操作。他们设计了一种多集群方法，可以处理各种工作负载和不同用户数下的大量增长数据。Khazaei等。 [145]，提出一个平台，对城市交通数据进行分析操作。该平台可供交通相关软件开发人员使用，也可由交通工程师和研究人员直接使用，以获取交通模式的见解。Chaolong等。 [146]研究了虚拟数据中心的发展趋势及其技术优势，提出了一种基于VMware vSphere的智能交通数据中心虚拟系统方案。大多伦多地区[147]提出了一个大数据模拟平台。该平台使大数据运输应用程序能够实时运行。

实时数据流处理功能是ITS中大数据处理平台的必要组成部分。因为有许多实时应用，如交通监控和公共交通时间表。基于传统的大数据处理系统，ITS中已经提出了大量的实时数据流系统。Guerreiro等。 [148]提出了一种用于智能交通系统的ETL（提取，转换和加载）架构，解决了高速公路动态收费的应用场景。所提出的架构能够使用大数据技术（如Spark on Hadoop和MongoDB）处理实时和历史数据。在[149]中提出了一种数据流处理平台，它支持共享多方数据源，软件组件甚至中间结果的机制。他们举了一个使用这个平台进行交通管理的例子。在[150]中提出了一种基于分布式计算平台的全面灵活的体系结构，用于实时交通控制。他们部分地在原型平台中实现了架构，该平台采用了Kafka，这是一种用于构建数据管道和流处理的最先进的大数据工具。

数据注入是大数据处理系统的另一个关键部分。它用于在大数据处理系统和关系数据库或大型机之间传输数据。作为一种流行的数据注入系统，Apache Sqoop已在ITS中得到广泛采用。例如，在[151]中，Sqoop与交通管理系统中的Hadoop一起使用。它还被用于处理车辆诊断数据并提供可供汽车生态系统中的参与者使用的有用结果[152]。在[153]中，Apache Sqoop用于摄取ITS关系数据。Apache Flume是另一种处理非结构化数据的流行数据注入系统，它已被用于处理ITS中的日志数据[154]。

# 七、开放的挑战

尽管大数据分析在ITS方面取得了巨大成就，但仍有大量的开放性挑战尚未得到充分研究。他们需要在未来的工作中得到解决。本节介绍在ITS中使用大数据分析的主要开放挑战如下。

•数据收集：由于车辆和行人频繁移动，在特定地点或特定时间运输中收集的数据可能不准确，不完整或不可靠。例如，并非所有车辆都嵌入了提供实时位置数据所需的技术，并且可能缺少来自道路传感器的道路交通数据。应对挑战者的一种可能方法是投资新的数据收集技术并提高数据收集能力。随着物联网的发展，每年都会发明新的传感器技术，这有助于提高数据收集和数据质量。此外，采用数据捕获自动化以最小化手动数据输入对于数据质量改进也是必不可少的。

•数据隐私：在大数据时代，最具挑战性和最关注的问题是隐私[155]。在数据传输，存储和使用过程中可能会泄露个人隐私[156]。从运输系统收集的数据曾经是非个人数据，例如车辆位置，交通流量数据。然而，由于公共和私营部门的个人数据收集随着时间的推移而增长，因此隐私问题一直受到关注例如，可以容易地收集个人和车辆的位置。如果这些数据没有受到严格保护，窃取这些数据的人将会损害数据所有者。因此，隐私保护对于ITS中的大数据应用来说是一件重要的事情。为防止未经授权披露个人隐私信息，政府应制定完整的数据隐私法，包括可以发布的数据，数据发布和使用的范围，数据分发的基本原则，数据可用性和其他方面[157]。运输部门应严格规范个人数据定义，加强数据安全认证管理，采用更先进的算法提高数据安全水平。

•数据存储：目前，数据量已从TB级别跃升至PB级别，数据存储容量的增长远远落后于数据增长。特别是在ITS中，它每天都会从各种传感器产生各种数据。传统的数据存储基础设施和数据库工具无法应对日益庞大和复杂的海量数据[158]。因此，设计最合理的数据存储架构已成为一项关键挑战。主要的公共云存储提供商（如Google和Microsoft）通过集成的大数据功能继续改进其服务，多云存储和混合存储正在成为大数据存储的关键领域。他们的计算突发功能在许多形式的计算密集型分析工作负载中具有优势。此外，将智能与存储相结合也是一个很好的解决方案。企业正在寻找能够在存储中提供集成分析的智能管理工具。这使他们能够进行资源监控并充分利用存储基础架构。

•数据处理：及时性对ITS中的大数据应用至关重要，这些应用包括交通数据预处理，交通状态识别，实时交通控制，动态路线引导和实时总线调度。包含来自不同来源的不同格式的交通数据必须与历史数据进行比较，然后在短时间内处理[159]。数据处理系统必须能够处理更复杂和不断扩展的数据。如何通过如此大而快的数据保证过程的及时性是一个巨大的挑战。最近出现了许多处理实时数据源的通用大数据框架，例如Apache Storm，Apache Flink，Apache Samza，Apache Spark Streaming和Kafka Streams。此外，还开发了专用于ITS的大数据处理框架，例如用于实时交通控制的平台，以及估算高速公路的平均速度和拥挤路段。这些处理框架为实时数据处理提供了良好的解决方案。

•数据开放：为了使运输服务用户和应用程序开发人员能够有效地发现和重新使用数据，需要将数据存档并以高质量公开访问。数据质量是指其准确性，完整性，可靠性和一致性[160]，[161]。如果没有良好的数据质量，大数据将会误导决策，甚至产生有害的结果。但是，打开高质量的数据可能需要时间和金钱。在以低成本快速打开数据和以高成本提供高质量数据之间存在权衡，这使得打开高质量数据成为一个更大的挑战。有效的解决方案包括采用自动数据捕获和/或利用人工智能来验证数据。此外，运输部门应制定数据管理流程，以确保原始和准确的数据。

# 八、结论

在本文中，我们介绍了大数据的发展和ITS的相关知识。讨论了在ITS中进行大数据分析的框架。我们总结了ITS中的数据源和收集方法，数据分析方法和平台以及大数据分析应用程序类别。我们在ITS中介绍了大数据分析的几种应用，包括资产维护，道路交通流量预测，道路交通事故分析，公共交通服务规划，个人旅行路线规划以及铁路运输管理和控制。本文讨论了在ITS中使用大数据分析的几个公开挑战，包括数据收集，数据隐私，数据存储，数据处理和数据打开。大数据分析将对智能交通系统的设计产生深远影响，使其更安全，更高效，更有效。

# 参考文献

[1] G. Bello-Orgaz, J. J. Jung, and D. Camacho, “Social big data: Recentachievements and new challenges,” Inf. Fusion, vol. 28, pp. 45–59,Mar. 2016.

[2] M. Chen, S. Mao, and Y . Liu, “Big data: A survey,” Mobile Netw.Appl., vol. 19, no. 2, pp. 171–209, Apr. 2014.

[3] H. Chen, R. H. Chiang, and V . C. Storey, “Business intelligence andanalytics: From big data to big impact,” MIS Quart., vol. 36, no. 4,pp. 1165–1188, 2012.

[4] T. B. Murdoch and A. S. Detsky, “The inevitable application of bigdata to health care,” JAMA, vol. 309, no. 13, pp. 1351–1352, 2013.

[5] M. Mayilvaganan and M. Sabitha, “A cloud-based architecture for big ata analytics in smart grid: A proposal,” in Proc. IEEE Int. Conf. Comput. Intell. Comput. Res. (ICCIC), Dec. 2013, pp. 1–4.

[6] L. Qi, “Research on intelligent transportation system technologies and applications,” in Proc. Workshop Power Electron. Intell. Transp. Syst., 2008, pp. 529–531.

[7] S.-H. An, B.-H. Lee, and D.-R. Shin, “A survey of intelligent transportation systems,” in Proc. Int. Conf. Comput. Intell., Jul. 2011, pp. 332–337.

[8] N.-E. El Faouzi, H. Leung, and A. Kurian, “Data fusion in intelligent transportation systems: Progress and challenges—A survey,” Inf. Fusion, vol. 12, no. 1, pp. 4–10, 2011.

[9] J. Zhang, F.-Y . Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, “Data riven intelligent transportation systems: A survey,” IEEE Trans. Intell. Transp. Syst., vol. 12, no. 4, pp. 1624–1639, Dec. 2011.

[10] Q. Shi and M. Abdel-Aty, “Big data applications in real-time traffic operation and safety monitoring and improvement on urban express-ways,” Transp. Res. C, Emerg. Technol., vol. 58, pp. 380–394, ep. 2015.

[11] N. Mohamed and J. Al-Jaroodi, “Real-time big data analytics: Applications and challenges,” in Proc. Int. Conf. High Perform. Comput. imulation, Jul. 2014, pp. 305–310.

[12] X. Lin, P. Wang, and B. Wu, “Log analysis in cloud computing environment with hadoop and spark,” in Proc. 5th IEEE Int. Conf. Broadband etw. Multimedia Technol. (IC-BNMT), Nov. 2013, pp. 273–276.

[13] M. Zaharia et al., “Fast and interactive analytics over Hadoop data with park,” USENIX Login, vol. 37, no. 4, pp. 45–51, 2012.

[14] D. Corrigan, P . Zikopoulos, K. Parasuraman, T. Deutsch, D. Deroos, nd J. Giles, Harness the Power of Big Data the IBM Big Data latform. 1st ed. New Y ork, NY , USA: McGraw-Hill, Nov. 2012.

[15] L. Basche, “Says solving ‘big data’ challenge involves more than just anaging volumes of data,” Bus. Wire, San Francisco, CA, USA, ech. Rep., Jun. 2011.

[16] M. Bagchi and P. R. White, “The potential of public transport smart ard data,” Transp. Policy, vol. 12, no. 5, pp. 464–474, 2005.

[17] M.-P. Pelletier, M. Trépanier, and C. Morency, “Smart card data use n public transit: A literature review,” Transp. Res. C, Emerg. Technol., vol. 19, no. 4, pp. 557–568, 2011.

[18] Y . Liu, X. Weng, J. Wan, X. Y ue, and H. Song, “Exploring data validity in transportation systems for smart cities,” IEEE Commun. Mag., vol. 55, no. 5, pp. 26–33, 2017.

[19] H. Nishiuchi, J. King, and T. Todoroki, “Spatial-temporal daily frequent trip pattern of public transport passengers using smart card data,” Int. J. Intell. Transp. Syst. Res., vol. 11, no. 1, pp. 1–10, 2013.

[20] M. A. Munizaga and C. Palma, “Estimation of a disaggregate multimodal public transport Origin–Destination matrix from passive smart card data from Santiago, Chile,” Transp. Res. C, Emerg. Technol., vol. 24, pp. 9–18, Oct. 2012.

[21] K. A. Chu and R. Chapleau, “Enriching archived smart card transaction data for transit demand modeling,” Transp. Res. Rec., J. Transp. Res. Board, pp. 63–72, Dec. 2008.

[22] M. A. Munizaga and C. Palma, “Estimation of a disaggregate multimodal public transport origin–destination matrix from passive smart card data from santiago, chile,” Transp. Res. C, Emerg. Technol., vol. 24, pp. 9–18, 2012.

[23] H. Gong, C. Chen, E. Bialostozky, and C. T. Lawson, “A GPS/GIS method for travel mode detection in New Y ork City,” Comput., Environ. Urban Syst., vol. 36, no. 2, pp. 131–139, 2012.

[24] X. Wang, S. Zhao, and L. Dong, “Research and application of traffic visualization based on vehicle GPS big data,” in Proc. Int. Conf. Intell. Transp., 2016, pp. 293–302.

[25] C. Asensio, J. López, R. Pagán, I. Pavón, and M. Ausejo, “GPS- based speed collection method for road traffic noise mapping,” Transp. Res. D, Transp. Environ., vol. 14, no. 5, pp. 360–366, 2009.

[26] J. C. Herrera, D. B. Work, R. Herring, X. Ban, Q. Jacobsond, and A. M. Bayen, “Evaluation of traffic data obtained via GPS-enabled mobile phones: The mobile century field experiment,” Transp. Res. C, Emerg. Technol., vol. 18, no. 4, pp. 568–583, Aug. 2010.

[27] K. G. Courage, M. Doctor, S. Maddula, and R. Surapaneni, “Video image detection for traffic surveillance and control,” Transp. Res. Center, Univ. Florida, Gainesville, FL, USA, Tech. Rep. TD100:FL96- 119, Mar. 1996.

[28] C. Grant, B. Gillis, and R. Guensler, “Collection of vehicle activity data by video detection for use in transportation planning,” J. Intell. Transp. Syst., vol. 5, no. 4, pp. 343–361, 2000.

[29] M. Kadaieaswaran, V . Arunprasath, and M. Karthika, “Big data solution for improving traffic management system with video processing,” Int. J. Eng. Sci., vol. 7, no. 2, 2017.

[30] J. Lopes, J. Bento, E. Huang, C. Antoniou, and M. Ben-Akiva, “Traffic and mobility data collection for real-time applications,” in Proc. IEEE Intell. Transp. Syst. (ITSC), Sep. 2010, pp. 216–223.

[31] C. Antoniou, R. Balakrishna, and H. Koutsopoulos, “Emerging data collection technologies and their impact on traffic management applications,” in Proc. 10th Int. Conf. Appl. Adv. Technol. Transp., A t h e n s , Greece, 2008.

[32] E. Huang, “Algorithmic and implementation aspects of on-line calibration of dynamic traffic assignment,” Ph.D. dissertation, Dept. Civil, Environ. Eng., Massachusetts Inst. Technol., Cambridge, MA, USA, 2010.

[33] E. Uhlemann, “Autonomous vehicles are connecting… [connected vehicles],” IEEE Veh. Technol. Mag., vol. 10, no. 2, pp. 22–25, Jun. 2015.

[34] C. Chen, T. H. Luan, X. Guan, N. Lu, and Y . Liu. (2017). “Connected vehicular transportation: Data analytics and traffic-dependent networking.” [Online]. Available: <https://arxiv.org/abs/1704.08125>

[35] X. Wang, C. Wang, J. Zhang, M. Zhou, and C. Jiang, “Improved rule installation for real-time query service in software-defined Internet of vehicles,” IEEE Trans. Intell. Transp. Syst., vol. 18, no. 2, pp. 225–235, Feb. 2017.

[36] J. Hu, L. Kong, W. Shu, and M.-Y . Wu, “Scheduling of connected autonomous vehicles on highway lanes,” in Proc. IEEE Global Commun. Conf. (GLOBECOM), Dec. 2012, pp. 5556–5561.

[37] R. King, “Traffic management in a connected or autonomous vehicle environment,” in Proc. Auto. Passenger Veh., May 2015, pp. 1–20.

[38] P. Bedi and V . Jindal, “Use of big data technology in vehicular ad-hoc networks,” in Proc. Int. Conf. Adv. Comput., Commun. Inform. (ICACCI), Sep. 2014, pp. 1677–1683.

[39] J. Contreras-Castillo, S. Zeadally, and J. A. G. Ibañez, “Solving vehicular ad hoc network challenges with big data solutions,” IET Netw., vol. 5, no. 4, pp. 81–84, Jul. 2016.

[40] M. Tan, I. W. Tsang, and L. Wang, “Towards ultrahigh dimensional feature selection for big data,” J. Mach. Learn. Res., vol. 15, pp. 1371–1429, Apr. 2014.

[41] A. Mahajan and A. Kaur, “Predictive urban traffic flow model using vehicular big data,” Indian J. Sci. Technol., vol. 9, no. 42, pp. 1–8, 2016.

[42] H. A. Najada and I. Mahgoub, “Anticipation and alert system of congestion and accidents in vanet using big data analysis for intelligent transportation systems,” in Proc. IEEE Symp. Ser. Comput. Intell. (SSCI), Dec. 2016, pp. 1–8.

[43] L. Gong, X. Liu, L. Wu, and Y . Liu, “Inferring trip purposes and uncovering travel patterns from taxi trajectory data,” Cartography Geograph. Inf. Sci., vol. 43, no. 2, pp. 103–114, 2016.

[44] C. Kang, Y . Liu, X. Ma, and L. Wu, “Towards estimating urban population distributions from mobile call data,” J. Urban Technol., vol. 19, no. 4, pp. 3–21, 2012.

[45] C. Chen, J. Ma, Y . Susilo, Y . Liu, and M. Wang, “The promises of big data and small data for travel behavior (aka human mobility) analysis,” Transp. Res. C, Emerg. Technol., vol. 68, pp. 285–299, Jul. 2016.

[46] J. Zeyu, Y . Shuiping, Z. Mingduan, C. Y ongqiang, and L. Yi, “Model study for intelligent transportation system with big data,” Proc. Comput. Sci., vol. 107, pp. 418–426, 2017.

[47] S. Liu, H. Cao, L. Li, and M. C. Zhou, “Predicting stay time of mobile users with contextual information,” IEEE Trans. Automat. Sci. Eng., vol. 10, no. 4, pp. 1026–1036, Oct. 2013.

[48] A. Gal-Tzur, S. M. Grant-Muller, T. Kuflik, E. Minkov, S. Nocera, and I. Shoor, “The potential of social media in delivering transport policy goals,” Transp. Policy, vol. 32, pp. 115–123, Mar. 2014.

[49] F. Alesiani, K. Gkiotsalitis, and R. Baldessari, “A probabilistic activity model for predicting the mobility patterns of homogeneous social groups based on social network data,” in Proc. 93rd Annu. Meeting Transp. Res. Board, 2014.

[50] Y . Chen, A. Frei, and H. Mahmassani, “From personal attitudes to public opinion: Information diffusion in social networks toward sustainable transportation,” Transp. Res. Rec., J. Transp. Res. Board, vol. 2, pp. 28–37, Nov. 2014.

[51] B. Pender, G. Currie, A. Delbosc, and N. Shiwakoti, “Social media use during unplanned transit network disruptions: A review of literature,” Transp. Rev., vol. 34, no. 4, pp. 501–521, 2014.

[52] X. Zheng et al., “Big data for social transportation,” IEEE Trans. Intell. Transp. Syst., vol. 17, no. 3, pp. 620–630, Mar. 2016.

[53] C. D. Cottrill and S. Derrible, “Leveraging big data for the development of transport sustainability indicators,” J. Urban Technol., vol. 22, no. 1, pp. 45–64, 2015.

[54] G. R. Grob, “Future transportation with smart grids & sustainable energy,” in Proc. 6th Int. Multi-Conf. Syst., Signals Devices (SSD), 2009, pp. 1–5.

[55] L. Zhu, F. R. Y u, B. Ning, and T. Tang, “Cross-layer handoff design in MIMO-enabled WLANs for communication-based train control (CBTC) systems,” IEEE J. Sel. Areas Commun., vol. 30, no. 4, pp. 719–728, May 2012.

[56] L. Zhu, F. R. Y u, B. Ning, and T. Tang, “Handoff performance improvements in MIMO-enabled communication-based train control systems,” IEEE Trans. Intell. Transp. Syst., vol. 13, no. 2, pp. 582–593, Jun. 2012.

[57] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, Machine Learning: An Artificial Intelligence Approach. N e w Y o r k , N Y , U S A : Springer, 2013.

[58] G. A. Seber and A. J. Lee, Linear Regression Analysis, vol. 936. Hoboken, NJ, USA: Wiley, 2012.

[59] H. Sun, H. Liu, H. Xiao, R. He, and B. Ran, “Use of local linear regression model for short-term traffic forecasting,” Transp. Res. Rec., J. Transp. Res. Board, pp. 143–150, Jan. 2003.

[60] Z. Shan, D. Zhao, and Y . Xia, “Urban road traffic speed estimation for missing probe vehicle data based on multiple linear regression model,” in Proc. 16th Int. IEEE Conf. Intell. Transp. Syst. (ITSC), Oct. 2013, pp. 118–123.

[61] N. Zenina and A. Borisov, “Regression analysis for transport trip generation evaluation,” Inf. Technol. Manage. Sci., vol. 16, no. 1, pp. 89–94, 2013.

[62] J. R. Quinlan, “Induction of decision trees,” Mach. Learn., vol. 1, no. 1, pp. 81–106, 1986.

[63] H. J. Payne and S. Tignor, “Freeway incident detection algorithms based on decision trees with states,” Transp. Res. Rec., vol. 682, pp. 30–37, Jan. 1978.

[64] J. AbelláN, G. LóPez, and J. De OñA, “Analysis of traffic accident severity using decision rules via decision trees,” Expert Syst. Appl., vol. 40, no. 15, pp. 6047–6054, 2013.

[65] C. Xie, J. Lu, and E. Parkany, “Work travel mode choice modeling with data mining: Decision trees and neural networks,” Transp. Res. Rec., J. Transp. Res. Board, pp. 50–61, Jan. 2003.

[66] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, “Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach,” Transp. Res. C, Emerg. Technol., vol. 13, no. 3, pp. 211–234, 2005.

[67] J. V an Lint, S. P . Hoogendoorn, and H. J. van Zuylen, “Accurate freeway travel time prediction with state-space neural networks under missing data,” Transp. Res. C, Emerg. Technol., vol. 13, nos. 5–6, pp. 347–369, Oct./Dec. 2005.

[68] X. Jin, R. L. Cheu, and D. Srinivasan, “Development and adaptation of constructive probabilistic neural network in freeway incident detection,” Transp. Res. C, Emerg. Technol., vol. 10, no. 2, pp. 121–147, 2002.

[69] X. Zhu, J. Guo, and W. Huang, “Short-term forecasting of remaining parking spaces in parking guidance systems,” in Proc. 95th Annu. Meeting Transp. Res. Board, 2016.

[70] L. Vanajakshi and L. R. Rilett, “Support vector machine technique for the short term prediction of travel time,” in Proc. IEEE Intell. V eh. Symp., Jun. 2007, pp. 600–605.

[71] Y . Bin, Y . Zhongzhen, and Y . Baozhen, “Bus arrival time prediction using support vector machines,” J. Intell. Transp. Syst., vol. 10, no. 4, pp. 151–158, 2006.

[72] J. Xiao and Y . Liu, “Traffic incident detection using multiple-kernel support vector machine,” Transp. Res. Rec., J. Transp. Res. Board, pp. 44–52, Dec. 2012.

[73] Y . Meng and X. Liu, “Application of K-means algorithm based on ant clustering algorithm in macroscopic planning of highway transportation hub,” in Proc. 1st IEEE Int. Symp. Inf. Technol. Appl. Edu. (ISITAE), Nov. 2007, pp. 483–488.

[74] R. P . D. Nath, H.-J. Lee, N. K. Chowdhury, and J.-W. Chang, “Modified k-means clustering for travel time prediction based on historical traffic data,” in Proc. Int. Conf. Knowl.-Based Intell. Inf. Eng. Syst., 2010, pp. 511–521.

[75] B. Abdulhai, R. Pringle, and G. J. Karakoulas, “Reinforcement learning for true adaptive traffic signal control,” J. Transp. Eng., vol. 129, no. 3, pp. 278–285, 2003.

[76] I. Arel, C. Liu, T. Urbanik, and A. G. Kohls, “Reinforcement learning based multi-agent system for network traffic signal control,” IET Intell. Transp. Syst., vol. 4, no. 2, pp. 128–135, 2010.

[77] A. L. C. Bazzan, “Opportunities for multiagent systems and multiagent reinforcement learning in traffic control,” Auto. Agents Multi-Agent Syst., vol. 18, no. 3, pp. 342–375, Jun. 2009.

[78] L. Li, Y . Lv, and F.-Y . Wang, “Traffic signal timing via deep rein forcement learning,” IEEE/CAA J. Automat. Sinica, vol. 3, no. 3, pp. 247–254, Apr. 2016.

[79] X. Ma, H. Y u, Y . Wang, and Y . Wang, “Large-scale transportation network congestion evolution prediction using deep learning theory,” PLoS ONE, vol. 10, no. 3, p. e0119044, 2015.

[80] H. Hu, B. Tang, X. Gong, W. Wei, and H. Wang, “Intelligent fault diagnosis of the high-speed train with big data based on deep neural networks,” IEEE Trans. Ind. Informat., vol. 13, no. 4, pp. 2106–2116, Aug. 2017.

[81] T. Chen, “Going deeper with convolutional neural network for intel ligent transportation,” Ph.D. dissertation, Dept. Elect. Comput. Eng., Worcester Polytech. Inst., Worcester, MA, USA, 2015.

[82] Y . Duan, Y . Lv, W. Kang, and Y . Zhao, “A deep learning based approach for traffic data imputation,” in Proc. IEEE 17th Int. Conf. Intell. Transp. Syst. (ITSC), Oct. 2014, pp. 912–917.

[83] N. Polson and V . Sokolov. (2016). “Deep learning for short-term traffic flow prediction.” [Online]. Available: https://arxiv.org/abs/1604.04527

[84] W. Huang, G. Song, H. Hong, and K. Xie, “Deep architecture for traffic flow prediction: Deep belief networks with multitask learning,” IEEE Trans. Intell. Transp. Syst., vol. 15, no. 5, pp. 2191–2201, Oct. 2014.

[85] Y . Lv, Y . Duan, W. Kang, Z. Li, and F.-Y . Wang, “Traffic flow prediction with big data: A deep learning approach,” IEEE Trans. Intell. Transp. Syst., vol. 16, no. 2, pp. 865–873, Apr. 2015.

[86] X. Ma, Z. Tao, Y . Wang, H. Y u, and Y . Wang, “Long short-term memory neural network for traffic speed prediction using remote microwave sensor data,” Transp. Res. C, Emerg. Technol., vol. 54, pp. 187–197, May 2015.

[87] J. Zhai, Y . Cao, and Y . Chen, “Semantic information retrieval based on fuzzy ontology for intelligent transportation systems,” in Proc. IEEE Int. Conf. Syst., Man Cybern. (SMC), Oct. 2008, pp. 2321–2326.

[88] S. Fernandez and T. Ito, “Driver behavior model based on ontology for intelligent transportation systems,” in Proc. IEEE 8th Int. Conf. Service-Oriented Comput. Appl. (SOCA), Oct. 2015, pp. 227–231.

[89] S. Fernandez and T. Ito, “Using SSN ontology for automatic traffic light settings on inteligent transportation systems,” in Proc. IEEE Int. Conf. Agents (ICA), Sep. 2016, pp. 106–107.

[90] D. Gregor et al., “A methodology for structured ontology construction applied to intelligent transportation systems,” Comput. Standards Interfaces, vol. 47, pp. 108–119, Aug. 2016.

[91] L. Zhao, R. Ichise, S. Mita, and Y . Sasaki, “Ontologies for advanced driver assistance systems,” J. Jpn. Soc. Artif. Intell., 2015, accessed: Aug. 12, 2016. [Online]. Available: <http://www.ei.sanken.osaka-> u.ac.jp/sigswo/papers/SIG-SWO-035/SIG-SWO-035-03.pdf

[92] D. Chen, F. Asplund, K. Östberg, E. Brezhniev, and V . Kharchenko, “Towards an ontology-based approach to safety management in coop- erative intelligent transportation systems,” in Proc. 10th Int. Conf. Depend. Complex Syst. Depcos-Relcomex, 2015, pp. 107–115.

[93] W.-D. Y ang and T. Wang, “The fusion model of intelligent transportation systems based on the urban traffic ontology,” Phys. Proc., vol. 25, no. 49, pp. 917–923, 2012.

[94] T. Toroyan, “Global status report on road safety,” Injury Prevention, vol. 15, no. 4, p. 286, 2009.

[95] T. F. Golob and W. W. Recker, “Relationships among urban freeway accidents, traffic flow, weather, and lighting conditions,” J. Transp. Eng., vol. 129, no. 4, pp. 342–353, 2003.

[96] G. Xiong, F. Zhu, H. Fan, X. Dong, W. Kang, and T. Teng, “Novel ITS based on space-air-ground collected big-data,” in Proc. IEEE Int. Conf. Intell. Transp. Syst., Oct. 2014, pp. 1509–1514.

[97] J. Lee and F. Mannering, “Impact of roadside features on the frequency and severity of run-off-roadway accidents: An empirical analysis,” Accident Anal. Prevention, vol. 34, no. 2, pp. 149–161, 2002.

[98] M. G. Karlaftis and I. Golias, “Effects of road geometry and traffic volumes on rural roadway accident rates,” Accident Anal. Prevention, vol. 34, no. 3, pp. 357–365, 2002.

[99] L.-Y . Chang and W.-C. Chen, “Data mining of tree-based models to analyze freeway accident frequency,” J. Safety Res., vol. 36, no. 4, pp. 365–375, 2005.

[100] M. Bédard, G. H. Guyatt, M. J. Stones, and J. P . Hirdes, “The independent contribution of driver, crash, and vehicle characteristics to driver fatalities,” Accident Anal. Prevention, vol. 34, no. 6, pp. 717–727, 2002.

[101] R. Li, C. Jiang, F. Zhu, and X. Chen, “Traffic flow data forecasting based on interval type-2 fuzzy sets theory,” IEEE/CAA J. Autom. Sinica, vol. 3, no. 2, pp. 141–148, Apr. 2016.

[102] D. Chen, “Research on traffic flow prediction in the big data environment based on the improved RBF neural network,” IEEE Trans. Ind. Informat., vol. 13, no. 4, pp. 2000–2008, Aug. 2017.

[103] S. Jeon and B. Hong, “Monte Carlo simulation-based traffic speed forecasting using historical big data,” Future Generat. Comput. Syst., vol. 65, pp. 182–195, Dec. 2016.

[104] X.-L. Liu, P . Jia, S.-H. Wu, and B. Y u, “Short-term traffic flow forecasting based on multi-dimensional parameters,” J. Transp. Syst. Eng. Inf. Technol., vol. 11, no. 4, pp. 140–146, 2011.

[105] H.-H. Dong, X.-L. Sun, L.-M. Jia, H.-J. Li, and Y . Qin, “Traffic condition estimation with pre-selection space time model,” J. Central South Univ., vol. 19, no. 1, pp. 206–212, 2012.

[106] M. Canaud, L. Mihaylova, J. Sau, and N.-E. El Faouzi, “Probabilty hypothesis density filtering for real-time traffic state estimation and prediction,” Netw. Heterogeneous Media, vol. 8, no. 3, pp. 825–842, 2013.

[107] T. L. Pan, A. Sumalee, R. X. Zhong, and N. Indra-Payoong, “Short-term traffic state prediction based on temporal–spatial correlation,” IEEE Trans. Intell. Transp. Syst., vol. 14, no. 3, pp. 1242–1254, Sep. 2013.

[108] C. Antoniou, H. N. Koutsopoulos, and G. Y annis, “Dynamic data driven local traffic state estimation and prediction,” Transp. Res. C, Emerg. Technol., vol. 34, pp. 89–107, Sep. 2013.

[109] B. Ghosh, B. Basu, and M. O’Mahony, “Multivariate short-term traffic flow forecasting using time-series analysis,” IEEE Trans. Intell. Transp. Syst., vol. 10, no. 2, pp. 246–254, Jun. 2009.

[110] J. Xu, D. Deng, U. Demiryurek, C. Shahabi, and M. van der Schaar, “Mining the situation: Spatiotemporal traffic prediction with big data,” IEEE J. Sel. Topics Signal Process., vol. 9, no. 4, pp. 702–715, Jun. 2015.

[111] H.-P . Lu, Z.-Y . Sun, and W.-C. Qu, “Big data and its applications in urban intelligent transportation system,” J. Transp. Syst. Eng. Inf. Technol., vol. 15, no. 5, pp. 45–52, 2015.

[112] C.-C. Lu, X. Zhou, and K. Zhang, “Dynamic origin–destination demand flow estimation under congested traffic conditions,” Transp. Res. C, Emerg. Technol., vol. 34, pp. 16–37, Sep. 2013.

[113] L. Alexander, S. Jiang, M. Murga, and M. C. González, “Origin destination trips by purpose and time of day inferred from mobile phone data,” Transp. Res. C, Emerg. Technol., vol. 58, pp. 240–250, Sep. 2015.

[114] J. B. Gordon, “Intermodal passenger flows London’s public transport network: Automated inference full passenger journeys using fare transaction vehicle-location data,” Ph.D. dissertation, Dept. Urban Studies Planning, Dept. Civil Environ. Eng., Massachusetts Inst. Technol., Cambridge, MA, USA, 2012.

[115] S. Tao, “Investigating the travel behaviour dynamics of bus rapid transit passengers,” Ph.D. dissertation, School Geograp., Planning Environ. Manage., Univ. Queensland, Brisbane, Qld, Australia, 2015.

[116] I. Gokasar and K. Simsek, “Using ‘Big data’ for analysis and improvement of public transportation systems in Istanbul,” Tech. Rep., 2014.

[117] J. Chan et al., “Rail transit OD matrix estimation and journey time reliability metrics using automated fare data,” PhD thesis, Dept. Civil Environ. Eng., Massachusetts Inst. Technol., Cambridge, MA, USA, 2007.

[118] J. L. Toole, S. Colak, B. Sturt, L. P . Alexander, A. Evsukoff, and M. C. González, “The path most traveled: Travel demand estimation using big data resources,” Transp. Res. C, Emerg. Technol., vol. 58, pp. 162–177, Sep. 2015.

[119] B. Ferris, K. Watkins, and A. Borning, “OneBusAway: Results fromproviding real-time arrival information for public transit,” in Proc.SIGCHI Conf. Human Factors Comput. Syst., 2010, pp. 1807–1816.

[120] [Online]. Available: http://inrix.com/mobile-apps/

[121] [Online]. Available: https://www.waze.com/

[122] [Online]. Available: http://moovitapp.com/

[123] [Online]. Available: http://gaode.com/

[124] “Opening up to open data, ” in Proc. Int. Assoc. Public Transp., 2014.

[125] B. Schultz, Ed., “Operational analytics keeps bay area trainson track,” All Analytics, May 2012. [Online]. Available: http://www.allanalytics.com/author.asp?section\_id=1411&doc\_id=244062

[126] M. Faizrahnemoon, A. Schlote, L. Maggi, E. Crisostomi, andR. Shorten, “A big-data model for multi-modal public transportationwith application to macroscopic control and optimisation,” Int. J.Control, vol. 88, no. 11, pp. 2354–2368, 2015.

[127] N. Van Oort, “Big data opportunities in public transport: Enhancingpublic transport by ITCS,” in Proc. IT-TRANS, Karlsruhe, Germany,Feb. 2014.

[128] Z. Jiang, C.-H. Hsu, D. Zhang, and X. Zou, “Evaluating rail transittimetable using big passengers’ data,” J. Comput. Syst. Sci., vol. 82,no. 1, pp. 144–155, 2016.

[129] J. Yin, D. Chen, and Y . Li, “Smart train operation algorithms basedon expert knowledge and ensemble CART for the electric locomotive,”Knowl.-Based Syst., vol. 92, pp. 78–91, Jan. 2016.

[130] D. Chen, T. Tang, C. Gao, and R. Mu, “Research on the error estimationmodels and online learning algorithms for train station parking in urbanrail transit,” China Railway Sci., vol. 31, no. 6, pp. 122–127, 2010.

[131] J. Zhou, “Applications of machine learning methods in problemof precise train stopping,” Comput. Eng. Appl., vol. 46, no. 25,pp. 226–230, 2010.ZHU et al.: BIG DA TA ANALYTICS IN ITSs: A SURVEY 397

[132] Z. Hou, Y . Wang, C. Yin, and T. Tong, “Terminal iterative learningcontrol based station stop control of a train,” Int. J. Control, vol. 84,no. 7, pp. 1263–1274, 2011.

[133] D. Chen, R. Chen, T. Tang, and Y . Li, “Online learning algorithms fortrain automatic stop control using precise location data of balises,”IEEE Trans. Intell. Transp. Syst., vol. 14, no. 3, pp. 1526–1535,Sep. 2013.

[134] A. Thaduri, D. Galar, and U. Kumar, “Railway assets: A potentialdomain for big data analytics,” Proc. Comput. Sci., vol. 53, no. 1,pp. 457–467, 2015.

[135] A. Núñez, J. Hendriks, Z. Li, B. De Schutter, and R. Dollevoet,“Facilitating maintenance decisions on the dutch railways using bigdata: The aba case study,” in Proc. IEEE Int. Conf. Big Data, Oct. 2014,pp. 48–53.

[136] J. Tutcher, “Ontology-driven data integration for railway asset mon-itoring applications,” in Proc. IEEE Int. Conf. Big Data, Oct. 2014,pp. 85–95.

[137] A. M. Zarembski, “Some examples of big data in railroad engineering,”in Proc. IEEE Int. Conf. Big Data, Oct. 2014, pp. 96–102.

[138] Asset Management Services, Network Rail, London, U.K., 2013.

[139] H. Li et al., “Improving rail network velocity: A machine learningapproach to predictive maintenance,” Transp. Res. C, Emerg. Technol.,vol. 45, pp. 17–26, Aug. 2014.

[140] F. Wang, T. Xu, T. Tang, M. Zhou, and H. Wang, “Bilevel featureextraction-based text mining for fault diagnosis of railway systems,”IEEE Trans. Intell. Transp. Syst., vol. 18, no. 1, pp. 49–58, Jan. 2017.

[141] X. Meng et al., “Mllib: Machine learning in apache spark,” J. Mach.Learn. Res., vol. 17, no. 34, pp. 1–7, 2016.

[142] R. Mian, H. Ghanbari, S. Zareian, M. Shtern, and M. Litoiu, “Adata platform for the highway traffic data,” in Proc. MESOCA, 2014,pp. 47–52.

[143] S. Zareian, R. V eleda, M. Litoiu, M. Shtern, H. Ghanbari, andM. Garg, “K-feed—A data-oriented approach to application perfor-mance management in cloud,” in Proc. IEEE 8th Int. Conf. CloudComput., Jun. 2015, pp. 1045–1048.

[144] M. Shtern, R. Mian, M. Litoiu, S. Zareian, H. Abdelgawad, andA. Tizghadam, “Towards a multi-cluster analytical engine for trans-portation data,” in Proc. Int. Conf. Cloud Auton. Comput. (ICCAC),2014, pp. 249–257.

[145] H. Khazaei, S. Zareian, R. Veleda, and M. Litoiu, “Sipresk: A big dataanalytic platform for smart transportation,” in Proc. EAI Int. Conf. BigData Anal. Smart Cities, 2015, pp. 419–430.

[146] J. Chaolong, H. Wang, and L. Wei, “Study of smart transportation datacenter virtualization based on vmware vsphere and parallel continuousquery algorithm over massive data streams,” Proc. Eng., vol. 137, no. 6,pp. 719–728, 2016.

[147] I. R. Kamel, H. Abdelgawad, and B. Abdulhai, “Transportation bigdata simulation platform for the greater toronto area (GTA),” in SmartCity 360°. New Y ork, NY , USA: Springer, 2016, pp. 443–454.

[148] G. Guerreiro, P . Figueiras, R. Silva, R. Costa, and R. Jardim-Goncalves,“An architecture for big data processing on intelligent transportationsystems: An application scenario on highway traffic flows,” in Proc.IEEE 8th Int. Conf. Intell. Syst. (IS), Sep. 2016, pp. 65–72.

[149] E. Bouillet et al., “Data stream processing infrastructure for intelligenttransport systems,” in Proc. IEEE 66th Veh. Technol. Conf. (VTC-Fall),Oct. 2007, pp. 1421–1425.

[150] S. Amini, I. Gerostathopoulos, and C. Prehofer, “Big data analyticsarchitecture for real-time traffic control,” in Proc. 5th IEEE Int. Conf.Models Technol. Intell. Transp. Syst. (MT-ITS), Jun. 2017, pp. 710–715.

[151] G. Zeng, “Application of big data in intelligent traffic system,” IOSRJ. Comput. Eng., vol. 17, no. 1, pp. 1–4, 2015.

[152] M. Tahmassebpour and A. M. Otaghvari, “Increase efficiency big datain intelligent transportation system with using IoT integration cloud,”J. Fundam. Appl. Sci., vol. 8, no. 3S, pp. 2443–2461, 2016.

[153] M. Chowdhury, A. Apon, and K. Dey, Data Analytics for IntelligentTransportation Systems. Amsterdam, The Netherlands: Elsevier, 2017.

[154] Z. Ji, I. Ganchev, M. O’Droma, L. Zhao, and X. Zhang, “A cloud-based car parking middleware for IoT-based smart cities: Design andimplementation,” Sensors, vol. 14, no. 12, pp. 22372–22393, 2014.

[155] M. Smith, C. Szongott, B. Henne, and G. von V oigt, “Big data privacyissues in public social media,” in Proc. IEEE Int. Conf. Digit. Ecosyst.Technol., Jun. 2012, pp. 1–6.

[156] Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, “Enabling publicauditability and data dynamics for storage security in cloud computing,”IEEE Trans. Parallel Distrib. Syst., vol. 22, no. 5, pp. 847–859,May 2011.

[157] O. Tene and J. Polonetsky, “Big data for all: Privacy and user control inthe age of analytics,” Northwestern J. Technol. Intell. Property, vol. 11,no. 5, 2012, Art. no. 1.

[158] M. Hilbert and P . López, “The world’s technological capacity to store,communicate, and compute information,” Sci., vol. 332, no. 6025,pp. 60–65, 2011.

[159] M. D. Assunção, R. N. Calheiros, S. Bianchi, M. A. S. Netto, andR. Buyya, “Big Data computing and clouds: Trends and future direc-tions,” J. Parallel Distrib. Comput., vols. 79–80, pp. 3–15, May 2013.

[160] J. Liu, J. Li, W. Li, and J. Wu, “Rethinking big data: A review onthe data quality and usage issues,” ISPRS J. Photogram. Remote Sens.,vol. 115, pp. 134–142, May 2016.

[161] J. Li and X. Liu, “An important aspect of big data: Data usability,”J. Comput. Res. Develop., vol. 50, no. 6, pp. 1147–1162, 2013.