

# 大数据在物联网中的应用

云 霜

## 一、 引言

大数据和物联网(IoT)正在高速发展, 并给各技术领域的应用带来了巨大的机遇, 包括: 智慧城市、智慧交通、智能电网以及远程医疗会诊等。由于物联网中通过不同传感器收集、处理数据, 因此, 广泛应用的物联网使得大数据分析具有挑战性。

国际数据公司(IDC)的报告显示, 到 2019 年, 大数据市场将超过 1250 亿美元。物联网大数据分析旨在帮助人们更好地理解数据, 从而做出更有效、明智的决定。大数据分析使数据挖掘者和科学家能够利用传统工具分析非结构化数据。此外, 大数据分析的目的是利用数据挖掘技术, 提取知识信息, 这些信息有助于进行预测, 识别趋势, 发现隐藏的信息, 并做出决定。

物联网数据不同于一般通过采集的大数据, 是因为其数据采集过程中涉及多种多样的传感器和物体, 从而采集的数据特性中包括: 异质性、噪声、多样性和快速增长性。物联网服务提供适当的资源和密集的平台应用程序, 以便各种已部署应用程序之间进行有效通信。这样的过程适合于满足物联网应用的要求, 并且可以解决大数据分析的一些问题。此外, 实施物联网和大数据集成解决方案有助于解决存储, 处理, 数据分析和可视化工具方面的问题, 还有助于提高智能城市中各种对象之间的协作和通信, 并且, 其他应用领域包括: 智能生态环境, 智能交通, 智能电网, 智能建筑和物流智能管理等也可从中收益。

## 二、 大数据以及物联网的概念

### 1、物联网

物联网可实现物物相连、物人相连、人与人之间任何时间、地点的有效连接。物联网使得传感器和设备可以在智能环境中无缝通信, 并以方便的方式实现跨平台的信息共享。

在过去几年中, 物联网已经成为一种新趋势, 移动设备, 交通设施, 公共设施和家用电器都可以用作物联网中的数据采集设备。日常生活中的电子设备(如: 手表)以及家用电器

（如冰箱）都可以连接到物联网网络，实现远程控制。物联网中，传感器设备感知并传送数据，设备和物体可以通过各种通信方式连接，例如，如蓝牙、WiFi、ZigBee 和 GSM，上述设备接收远程控制设备的命令，从而实现了计算机的系统与物理世界的集成，以提高人们的生活水平。

## 2、大数据

由传感器、设备、社交、媒体、医疗保健应用程序、温度传感器以及各种其他软件应用程序和数字设备不断生成的结构化、非结构化或半结构化数据的数据量正在持续增长。此大规模的数据产生了“大数据”。传统的数据库系统在存储、处理和分析快速增长的数据量时效率低下。

麦肯锡全球研究院将大数据定义为：一种规模大到获取、存储、管理、分析等方面打打超出传统数据库工具处理能力范围的数据集[1]。“数字宇宙”将大数据技术定义为新一代的技术和架构，旨在从海量数据中提取各种各样有价值的信息，通过高速捕获、发现和分析步骤。换句话说，大数据就是能够被记录、采集和开发利用的实时有效的、数据庞大的、多样化的数据集、数据流和数据体[1]。大数据具有以下性质：（1）数据量庞大；（2）价值密度低，获取有用信息，难度大；（3）多样数据形式：结构化、半结构化、非结构化组成；（4）实时性[1]。

大数据具有 3V 特点：大规模（Volume）、多样性（Variety）、高速性（Velocity）。随着大数据研究的深入，参见表 1，大数据的扩展为 6v 特点：大规模（Volume）、多样性（Variety）、高速性（Velocity）、价值(Value)、真实性（Veracity）、易受攻击(Vulnerable) [2]。

特性	解释
大规模（Volume）	（1）大数据规模庞大、 （2）大数据的数据来源广泛 （3）数据常常以分布式存在
多样性（Variety）	大数据的来源多、种类多样、数据和分析方法多样
高速性（Velocity）	在线或实时分析，要求短时间对海量数据进行快速获取、处理、分析和报告
价值(Value)	价值密度低，一旦从海量数据分析获取所需信息，其价值往往也会非常高
真实性（Veracity）	（1）大数据整个生命周期保证是真实的，它的数据源可信、处理过程建立在可信计算基础上，存储受到保护且数据真实可信； （2）大数据使用过程中应该是经过授权和访问控制的，它

	的修改能够溯源
易受攻击(Vulnerable)	(1) 大数据分析结构价值高，容易刺激攻击者趋利本质而发动攻击； (2) 大数据发展不够成熟，在数据获取、存储、处理、管理等方面还存在许多漏洞，易受到攻击 (3) 攻击者可以利用大数据发动大规模持续攻击，破坏性更大

表 1 大数据 “6V” 特性具体描述[2]

### 三、 大数据相关的技术方法

根据物联网应用的要求使用不同的分析类型。 参见表 2，本小节将从实时，离线，内存级别，商业智能（BI）级别和大规模级别分析类别中讨论这些分析类型。

#### 1、 现有的分析系统

**Real-time 实时：**实时分析通常是对从传感器收集的数据执行的。在这种情况下，数据不断变化，因此，需要快速的数据分析技术以在短时间内获得分析结果。

**Off-line 离线：**当不需要快速响应时可使用离线分析系统。 例如，许多互联网企业使用基于 Hadoop 的离线分析架构来降低数据格式转换的成本。此类分析可提高数据采集效率。

**Memory-level 内存级别：**在数据的大小小于集群的内存时应用。到目前为止，集群的记忆已达到太字节（TB）水平。因此，需要内部数据库技术来提高分析效率。内存级分析适合进行实时分析。

**BI-level 商业智能级别：**当数据的大小大于内存级别时，将采用 BI 分析。BI analytic 目前支持太字节级数据。此外，BI 有助于从大量数据中发现战略商业机会，发现新的机会并实施有效的战略，提供有竞争的市场优势。

**Massive-level:** 当数据量超过 BI 分析产品和传统数据库的全部容量时，就需要进行大规模的分析。大规模分析通过从数据中提取有价值的信息，从而有助于创建商业基础，并提高市场竞争力。此外，大规模的分析可以获得准确的数据，降低业务决策涉及的风险。此外，大规模的分析还提供了有效的服务。

分析类型/级别	使用方式	现有架构/工具	优势/使用归属类型
实时	分析从传感器收集的数据	Greenplum HANA	并行执行

离线	当不需要快速响应时	Scribe Kafka Timetunnel Chukwa	高效数据采集 效率 降低成本
内存级别	在数据的大小小于集群的内存时应用	MongoDB	实时分析
商业智能级别	当数据的大小大于内存级别时	数据分析方法	实时和离线
大规模级别	当数据量超过 BI 分析产品和传统数据库的全部容量时	MapReduce	离线

表 2 大数据相关的技术方法

## 2、大数据与物联网之间的关系

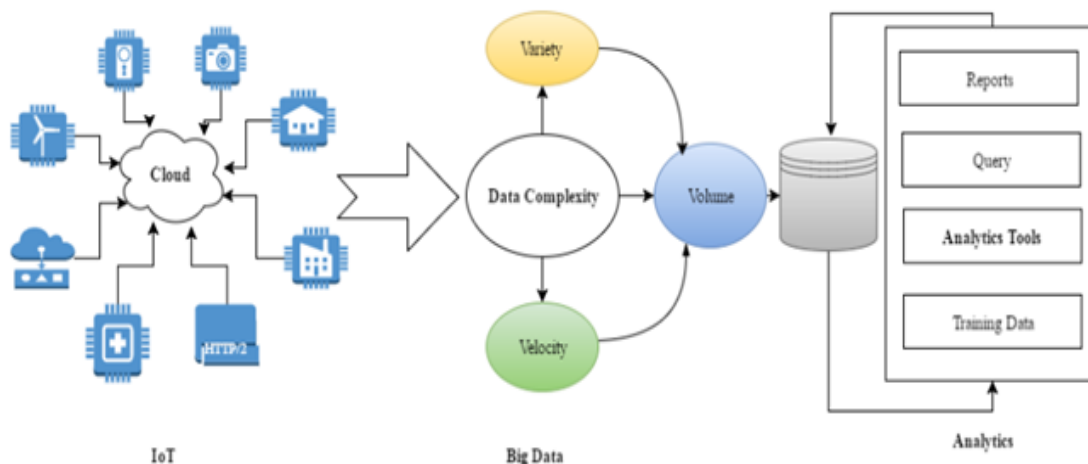


图 1 大数据与物联网之间的关系[3]

大数据分析正迅速成为物联网中的关键技术。物联网中的大数据分析需要实时处理大量数据，并将数据存储在各种存储技术中。考虑到大部分非结构化数据都是直接从支持 web 的“物体”中收集的，大数据分析需要执行如闪电般的快速分析，以允许企业和组织快速观察、快速决策，并与人和其他设备进行交互。

物联网和大数据之间的关系如图 1 所示，可以分为三个步骤来实现物联网数据的管理。第一步为管理物联网数据源，其中连接的传感器设备通过应用程序来实现交互。例如，诸如 CCTV 摄像机、智能交通灯和智能家居设备之类的设备的交互，产生了大量不同格式的数据源，这些数据可以存储在云上。在第二步中，生成的数据称为“大数据”，之所以叫做“大数据”，这是因为它们的速度、容量和多类型。海量的数据共享在分布式可容错的数据库中。

最后一步是应用分析工具，如 MapReduce，Spark，Splunk 和 Skytree，上述工具可以分析存储的物联网数据集。4 个层级的分析从训练数据开始，然后转向至分析工具，查询和报告。

### 3、大数据分析的方法

参见图 2,大数据分析可分为 4 个类别：分类、聚类、关联规则挖掘和预测。

分类是一种利用先前的知识作为训练数据对数据对象进行分组的监督学习方法。一个预定义的类别被分配给一个对象，从而对对象的组或类进行预测(见图 2)。对于物联网大数据来说，寻找未知或隐藏的数据更具有挑战性。

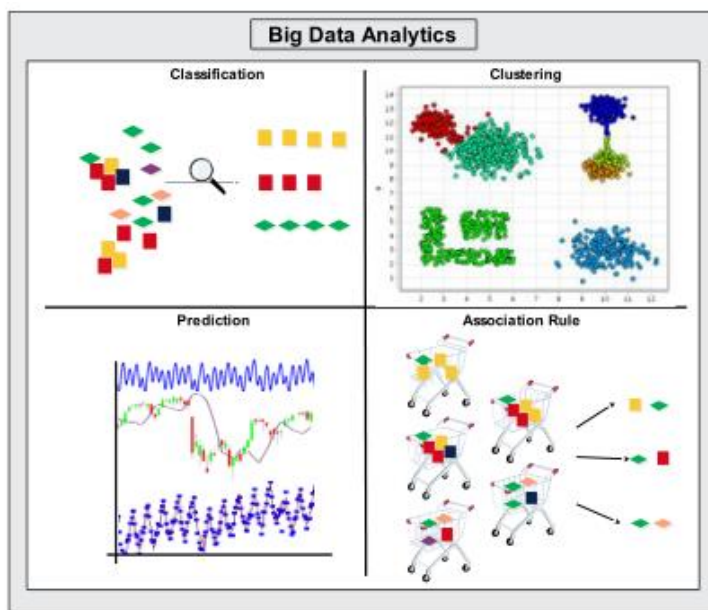


图 2 大数据分析的方法[3]

贝叶斯网络是一种提供模型可解释性的分类方法。相对于传统的结构化数据格式，贝叶斯网络更适合分析复杂数据结构。SVM 支持向量机是大数据分析的一种分类方法，可以有效地分析数据模式和创建组。SVM 利用统计学习理论分析数据模式并创建组。SVM 支持向量机在大数据分析中的应用包括：文本分类、模式匹配、健康诊断和商业。类似地，KNN 是为了一种能够从大数据集中发现隐藏模式的有效方法。改进的 KNN 算法，应用于异常检测、高维数据、科学实验。

聚类是大数据分析方法的另一种数据挖掘技术。与分类相反，聚类使用无监督学习方法，并根据特定特征为给定对象创建组。正如图 2 中所呈现的那样，以簇的形式对大量对象进行分组使得数据操作变得简单。众所周知的聚类方法是层次聚类和分区。

市场分析和商业决策是大数据分析中最重要的应用。关联规则挖掘是一种数据挖掘算法，它包括：识别不同对象、事件或其他实体之间的关系，以分析市场趋势,消费者购买行为和产品需求预测(见图 2)。关联规则挖掘侧重于根据数字和非数字数据的出现频率，识别和创建规则。根据关联规则，数据处理可以两种方式执行。顺序数据处理方法使用基于优先级的算法(如 MSPS[59]和 LAPIN-SPAM[60])来识别关联关系。另一种重要的数据处理方法是时间序列分析，它使用算法分析连续数据中的事件模式。

预测分析使用历史数据（称为训练数据）确定数据中的趋势或行为。 SVM 和模糊逻辑算法用于识别独立变量和因变量之间的关系，并获得预测的回归曲线，例如自然灾害。此外，可通过预测分析分析客户购买预测和社交媒体趋势。

### 4、实现的大数据分析的物联网架构

图 3 展示了大数据分析的物联网架构。在该图中，传感器层包含所有通过无线网络连接的传感器设备和对象。这些无线网络通信可以是 RFID、WiFi、超宽带、ZigBee 和蓝牙。物联网网关允许互联网和各种网站的通信。上层涉及大数据分析，其中从传感器接收的大量数据存储在云中并通过大数据分析应用程序访问。这些应用程序包含 API 管理和一个仪表板以助于与处理引擎的交互。

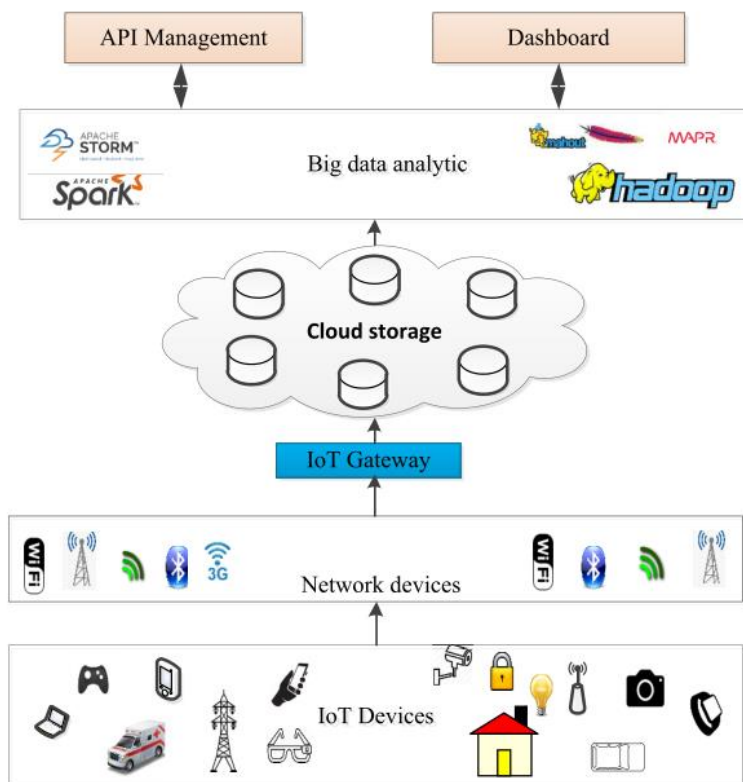


图 3 大数据分析的物联网架构[3]

## 四、 大数据在物联网中的应用情况

### 1、智能计量

智能计量是物联网应用之一，它可以从不同的来源生成大量数据，例如智能电网、水箱水位、水流。一个智能电表能够以电子方式记录消耗的电量。在物联网环境下收集和分析智能电表数据，在很大程度上可以帮助电网公司预测用户用电量。此外，智能电表的分析还可以用来预测需求以防止危机和通过具体的定价计划满足战略目标。

### 2、智能交通

智能交通系统是物联网应用之一，旨在支持智能城市。传统的基于图像处理的交通系统容易受到暴雨、浓雾等天气条件的影响，从而导致采集的图像不清晰。智能交通系统旨在为智能城市的管理部署强大而先进的通信技术。此外，将物联网引入交通领域将减少交通拥堵。这项技术可以改善现有的交通系统的性能，使车辆能够在没有人为干预的情况下，以系统的方式有效地相互沟通。

卫星导航系统和传感器也可以应用于卡车、轮船和飞机上。可以通过使用大量公共数据来优化公共交通的路线，这些数据可以是，例如交通堵塞、道路状况、送货地址、天气状况和加油站。例如，在行驶路线更改时，可以优化、重新计算更新的信息 (路线、成本)，并将结果实时地传递给司机。这些车辆内置的传感器还可以提供实时信息来检测发动机的健康状况，确定设备是否需要维护，并预测故障。

### 3、智能供应链

嵌入式传感器技术可以支持双向通信，并为全球超过 100 万部电梯提供远程接入。现场和场外的技术人员使用采集的数据用于运行诊断和修复选项，以做出适当的决策，保障机器正常运行并提高用户服务质量。大物联网数据分析允许供应链执行决策和控制外部环境。启用 IOT 的工厂的设备之间通过数据参数 (机器设置、温度) 进行通信，通过改变设备设置和流程 workflow 来优化性能。在途可见性是另一个用例，它将在未来的供应链中发挥至关重要的作用。过境可见性的关键技术是 RFID 和基于云的全球定位系统(GPS)，它提供定位、身份和其他跟踪信息。这些数据将成为物联网技术支持的供应链的支柱。通过 RFID 和 GPS 技术收集的数据将允许供应通过预测到货时间，提高自动化发货和准确的发货信息。类似地，管理

人员将能够监视其他信息，例如温度控制。

### 4、智能农业

智能农业是大物联网数据分析的一大应用。传感器安装在田间，获取土壤水分水平、植物树干直径、微气候条件、湿度水平等数据，以进行天气预报。传感器使用网络和通信设备传输获得的数据，这些数据通过物联网网关和网络到达分析层，分析层处理从传感器网络获得的数据以发出命令，从而实现自动控制气候，及时控制灌溉和湿度控制。

### 5、智能电网

智能电网是新一代的电网，利用双向通信技术和计算能力实现电力管理以及分配，并进行升级以提高可靠性、安全性、实时控制和监控的效率。电力系统的主要挑战之一是整合可再生能源和分散的能源。电力系统需要一个智能电网来管理不稳定的行为分布式能源。传感器和设备持续快速地生成与控制回路和保护相关的数据，且需要实时处理和分析机器对机器(M2M)或人机(HMI)交互的信息，以便向系统发出控制命令。

### 6、智能交通灯系统

智能交通灯系统由节点组成，节点与物联网传感器和设备进行本地交互，以检测车辆、骑车者和行人。这些节点利用邻近的交通灯来检测临近的交通工具的速度和距离，管理交通信号。此外,智能交通灯系统可以将收集到的物联网数据发送到云存储中进行进一步分析。

## 五、 大数据在物联网中应用的发展趋势

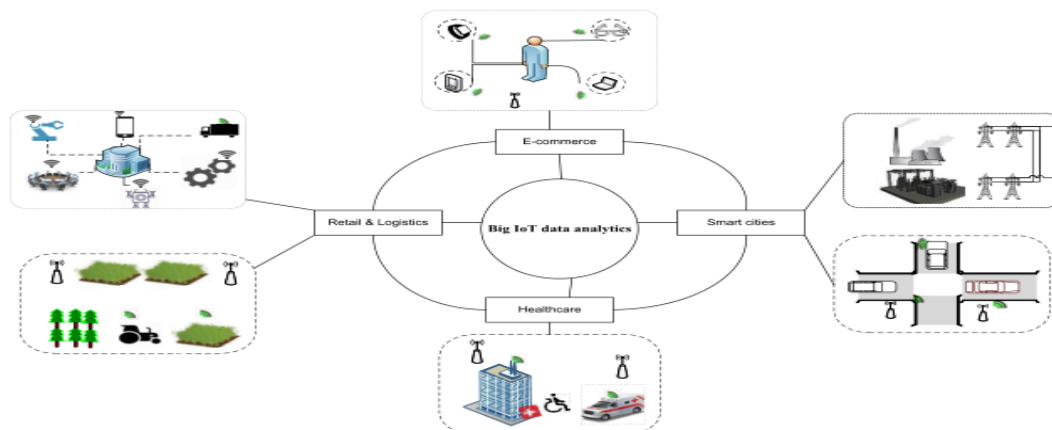


图 4 大数据在物联网中应用的发展趋势[3]



## 1、电子商务

大数据与物联网的融合为构建智能环境带来了新的挑战 and 机遇。大物联网数据分析在几乎每个行业都有广泛的应用。然而，其应用成功的主要领域是电子商务、收益增长、客户规模的扩大、销售预测结果的准确性、产品优化、风险管理和客户分类。

## 2、智慧城市

大数据为智慧城市提供了新的机会，各种设备在智能环境中连接到网络中并共享信息。此外，在云计算技术发展，存储数据的成本已经大大降低。因此，大数据在智能城市中的作用可能会涉及经济的每个领域。

## 3、零售业和物流领域

作为一项新兴技术，物联网有望在零售和物流领域中发挥关键作用。然而，物联网设备每天都会产生大量数据。因此,强大的大数据分析使得企业能够从物联网技术产生的海量数据中获得有价值的信息。将数据分析应用于物流数据集可以改善客户的发货体验。此外，零售企业可以通过分析客户数据来获得额外的利润，这些数据可以预测商品的趋势和需求。通过查看客户数据，可以有效地规划价格计划和实现季节性促销，以实现利润最大化。

## 4、健康

近年来，智能健康监测设备出现了巨大的增长，这些设备可产生大量的数据。因此，将数据分析应用于从胎儿监护仪、心电图、体温监护仪或血糖水平监护仪收集的数据，可以帮助医疗专家有效地评估患者的身体状况。此外，数据分析使医疗人员能够在早期诊断严重疾病，以挽救生命。此外，数据分析提高了临床护理质量，确保了患者的安全。

# 六、 大数据在物联网中未来的挑战

## 1、隐私

尽管数据来自于匿名用户，但系统在使用大数据分析工具进行推断或恢复个人信息仍然会出现隐私问题。随着大数据分析技术的发展，隐私问题已成为数据挖掘领域的核心问题。因此，必须保护用户的敏感信息并防止外部窃取。虽然匿名和加密提供了加强保护隐私的方

法，但使用者使用时需考虑道德标准，例如使用什么，如何使用以及为什么使用。

与 IoT 数据相关的另一个安全风险是所使用的设备类型的异构性和生成的数据的性质，例如原始设备，数据类型和通信协议。这种异构的物联网架构对安全人员来说是新的，因此会导致风险。因此，任何攻击都会危及新系统的安全性并导致互连设备的断开。

通过物联网生成的数据，可能会出现以下安全问题:( a ) 很难及时更新系统，( b ) 识别合法流量模式中的可疑流量模式以及可能无法捕获不明信息，( c ) 专有的和特定于供应商的程序很难发现隐藏的或零天攻击 ( d ) 和协议融合方面造成困难 - 尽管 IPv6 目前与最新规范兼容，该协议尚未完全部署。因此，在 IPv4 上应用安全规则可能不适用于保护 IPv6。

## 2、数据挖掘

数据读写:大容量、高速度、高多样性的物联网数据对探索、集成、异构通信和提取过程提出了挑战。数据的大小和异构性带来了新的数据挖掘需求和数据的多样性来源也带来了一个挑战。此外，与小数据集相比，大数据集包含更多的异常和歧义，需要额外的预处理步骤，如清理、减少和传输。另一个问题是如何从大量不同的数据中提取准确和有价值的信息。因此，从复杂数据中获取准确的信息需要分析数据属性和发现不同数据点之间的关联。

研究人员引入了并行和顺序编程模型，并提出了不同的算法，以最大限度地减少查询响应时间，同时处理大数据。此外，研究人员以不同的方式选择现有的数据挖掘算法 ( a ) 改进单一来源发现，( b ) 实现多源平台的数据挖掘方法，( c ) 研究和分析动态数据挖掘方法和流数据。因此，引入了并行 k 均值算法和并行关联规则挖掘方法。但是，在并行计算中可能发生同步问题，同时信息交换还发在不同的数据挖掘方法中。这一瓶颈已成为大型物联网数据分析中应该解决但悬而未决的问题。

## 3、可视化

可视化是大数据分析中的一个重要实体，尤其是在数据生成量巨大的物联网系统时。此外，由于大数据规模大、维数高，数据可视化的实现难度较大。因此，大数据分析和可视化应该无缝结合，以获得大数据结合物联网应用的最佳结果。然而,在异构和不同数据(非结构化、结构化和半结构化)的情况下进行可视化是一项具有挑战性的任务。设计与大数据索引框架兼容的可视化解决方案是一项艰巨的任务。因此，可以部署支持 GUI 设施的云计算体系结构，以便更好地了解物联网大数据趋势。

虽然引入了不同的降维方法，但是，这些方法不适用于所有类型的呈现数据。此外，由于功率和带宽的限制，数据应该保存在本地，以便有效地获取可用信息。此外，在物联网环境中，可视化软件应以参考位置的概念运行，以实现高效的结果。随着物联网数据量的快速增长，对海量并行化的需求是可视化领域的一项具有挑战性的任务。

#### 参考文献

- [1] 周兴元，浅谈大数据背景下的物联网发展，信息化建设, 2018,9.20
- [2] 吕登龙，朱诗兵，大数据及其体系架构与关键技术，装备学院学报, 2017,2 (28): 87-95
- [3] MOHSEN MARJANI, FARIZA NASARUDDIN, ABDULLAN GANI, AHMAD KARIM, IBRAHIM ABAKER TARGIO HASHEM, AISHA SIDDIQA, AND IBRAR YAQOOB , Big IoT Data Analytics: Architecutre, Opportunities, and Open Research Challenges, IEEE Access, 2017, volume 5, 5247-5259



临菲信息技术港



临菲信息技术港 公众号



临菲学堂