

物联网数据质量研究报告

摘要：物联网的发展给予了大量非传统设备接入网络的可能，这些非传统设备有以下两个特点：规模大且种类多，同时，这些设备产生的数据也有规模庞大、数据结构复杂的特点。大量非传统设备产生的海量异构数据给现有网络带来了新的问题挑战——如何评估海量异构数据的质量、数据质量带来哪些问题及怎样保证数据质量成为物联网发展亟待解决的问题。本文调研了当前物联网有关数据质量问题的研究现状，针对上述三个问题，分别围绕物联网数据质量评估方法、数据质量问题以及解决数据质量问题的方法做了阐述总结，最后对未来的研究方向做了展望。

关键词：数据质量、物联网、异构数据

1 引言

物联网（Internet of Things, IOT）是一种将日常万物接入互联网的新型网络结构，不同于现有网络中广泛存在的手持终端，物联网更加偏重于非传统的低成本硬件，如传感器、频标签等。通过大量上述非传统设备接入现有网络，实现物物、物人的广泛连接，进而实现对接入网络的各类资源的智能化感知、交互与管理。从实体来看，最早的物联网实例始于射频识别（RFID）^[1]，目前该技术已被广泛应用于物流交通、医疗健康等多个行业；从概念来看，由最早的 RFID 技术发展而来的与物的连接是人们看得见的最早的物联网网络，正因如此，Ghallab、Fahmy 等人将物联网定义为“一组通过射频识别连接的智能事物”^[2]，由此可看出射频识别技术对物联网的重要意义；由数据出发，Karkouch 等人将物联网描述为数据的主要消费者是智能设备的网络^[3]。总而言之，物联网的最终目标实现真正的“万物互联”，其实体构成也应该是能够容纳参差不齐的各类硬件设备的新型网络。

1.1 研究背景

尽管物联网不同于现有网络，但并不意味着物联网是完全脱离现有网络的新型实体，当前物联网的发展建立在现有传统网络的基础之上。由此意味着大量不同于以往的设备加入现有网络，这些非传统设备的加入不但使网络的边缘设备构成发生了很大变化，而且也改变了网络中的数据构成，前者使数据的生产者发生变化，由以传统的、较为统一智能设备为主的数据生产者变为数量庞大、种类复杂的各类新型硬件为主要的的数据生产者；后者意味着网络中的数据量变大，且数

据变得更加多元复杂。同时，由于这些非传统设备往往是低成本的硬件或传感器，其算力、存储力有限，因此数据时刻需要上载，需要一定的实时性，且由于硬件成本导致设备的纠错能力受限、上载数据量大都将导致产生数据中的冗余量、错误量更大。

以上不难发现，由物联网硬件特点带来的几点问题皆与数据有关，更确切地说，与数据质量有关。无论是传统互联网还是物联网，网络结构的出现都可以视为服务于数据的连接实体，整个网络可以理解为数据生产者、数据消费者、以及数据传输等中间环节的整合。Karkouch 等人认为，数据是物联网中最有价值的资产，因为它们为应用程序提供了有关智能服务的特定实体、现象描述或个人见解^[3]，指出了数据的重要性；此外，许多研究也阐明了数据质量对数据挖掘过程的重要性以及低数据质量对结果可靠性的影响^[4]，由此可见，数据质量无论对于传统的网络服务，或是对数据挖掘等前沿领域都是及其重要的，数据质量很大程度上影响决策的正确性、服务的可靠性及网络的稳定性。

遗憾的是，在传统网络中，由于智能设备算力较高、应用场景较为单一等原因，数据能够在生产端得到充分的处理，上载到网络的数据格式较为一致且质量较为有保证；而物联网中众多非传统硬件设备的加入改变了上述局面，使得物联网数据较之前有上述数据量大、结构复杂、冗余及错误数据更多的特点。这些特点很大程度上影响了物联网数据质量，因此，无论对于物联网本身的发展还是对传统网络的安全性而言，物联网的大量异构数据质量应该得到保证，由此，数据质量已成为一个亟待解决的关键问题，本文主要围绕数据质量问题，调研讨论数据质量的评价指标及现有的解决方法。

1.2 问题阐述

值得注意的是，数据质量和数据质量问题是不同的，简言之，前者指数据的好坏，后者关乎数据好坏带来的后果。此外，针对数据质量问题就会有相应的解决方法。本文也正是围绕上述三个概念展开，即：数据质量、数据质量问题及如何解决数据质量问题。围绕数据质量，主要的问题并不在于数据质量的概念定义，而在于更具有实际意义的问题——如何量化评估数据质量。围绕数据质量问题，本文主要对数据质量问题做了区别与数据质量的阐述，并对文献中出现的数据质量问题做了总结；针对数据质量问题的解决方法，本文整理阐述了当前的解决方案并对其做了简要分析。总结起来，可以将数据质量相关问题总结为以下三个命题，即本文的主要贡献在于：

1. 如何评估数据质量；
2. 数据质量带来哪些问题；
3. 数据质量问题的解决方法有哪些；

针对上述问题，本文通过关键词搜索对近年文献做了调研，在后续第二章节分别做了解答，在本文最后的第三部分，对三个问题做了总结并对未来的数据质量相关研究做了展望。

2 研究现状

本章主要阐述当前与数据质量相关的研究工作，首先说明数据质量评估与数据质量问题的主要区别，后续主要分三小节：数据质量评估、数据质量问题及数据质量问题的解决方法，分别作为上文三个问题的主要解答。数据质量评估主要总结文献中出现的数据质量评估维度，下面介绍数据质量评估维度与数据质量问题的主要区别。

文献^[1]首次将数据质量评价维度与数据质量问题区分开，但文中并没有给出更多关于二者区别的阐述，而是罗列现有文献中出现的概念。本文同样认为，将数据质量评估维度与数据质量问题混为一谈是不合适的，很多文章都忽略了这一点，或是将二者混为一谈，或是将数据质量问题错误地作为评价指标使用，本文认为有必要对二者进行区分。

正如上述对于数据质量和数据质量问题的阐述——前者是数据的好坏，后者是其后果一样，本文对于二者区分的阐述是，对于数据质量评价维度，应当更多地从数据本身阐述，其评价维度应当是客观的、通用性更强的，例如精确性、完整性、时效性等；而对于数据质量问题，则侧重于保证上述基本数据质量因素的前提下，对于使用者来说还需考虑的问题，更多地从数据的语义考虑，数据质量问题是针对性更强的、更难以量化的，例如：数据冗余、上下文错误、噪声等。总结起来，二者主要在以下几点区分：

1. 数据质量评估维度是通用的指标，数据质量问题是描述在保证数据质量的前提下对不同应用场景的利用好坏。前者是对数据本身的客观评价，后者是应用于不同场景下的主观问题；
2. 数据质量的评估维度更容易量化评价，数据质量问题往往是在应用语义上更难以量化的部分；

区分数据质量评估维度和数据质量问题的目的在于，具体应用场景能够根据数据质量的评价维度量化质量好坏，而对数据质量问题做有针对性的优化。例如，根据评价维度判定基础数据的可靠性后，根据不同应用场景要求对数据质量问题之一的冗余度做针对性优化解决；反之，若判定数据不可靠后，后续的所有操作都是不必要的了——冗余度的判定不是通用的，不同应用场景对冗余的判定有所不同，对冗余度的要求也有所不同，而无论什么场景，对数据可靠性的要求是相同的，这便是划分数据质量评估维度与数据质量问题的目的所在。

2.1 数据质量评估

当前工作中针对数据质量的评估方法有很多，但大致可分为两类，一类是从数据消费者的角度出发，认为数据应当充分贴合使用场景，相应地，数据质量也被定义为“数据对于分析目的的充分性”以及“数据满足消费者的程度”^[3]；另一类是从更加普适性的评价指标出发，试图给出通用的量化评价标准。不难发现，两种方法对数据质量的量化方法并无根本区别，主要不同在于评价方法在针对性与通用性方向上的取舍，基于上述分析，本文从普适性的角度出发，将文献中出现的数据质量评价指标做汇总，其目的并不是罗列作者的计算公式，而是阐述一般的评估数据质量的维度。

文献^[1]同样对数据质量的评价维度做了详细的工作，并对文献中存在相似定义的出现次数做了统计，总结为以下表 2.1：

表 2.1 数据质量评价维度汇总

维度	描述	出现次数
Accuracy	准确性衡量数据的捕获值与其原始值的接近程度	26
Precision	-	3
Usability	可用性指在数据失去价值之前可以维护的时间	1
Relevance	相关性表明获得的数据对应用程序有多大价值	3
Ease of Understanding	-	1
Privacy	隐私性表明对被允许访问数据的人的限制	13
Objectivity	-	1
Reputation	-	1
Granularity	粒度是指衡量存储数据详细程度的空间和时间维度	1
Integrity	-	1
Currency	用户对最新数据的感知程度	2
Completeness	完整性定义了数据集中可用值的范围	18
Timeliness	及时性、延迟、流通性或波动性显示捕获数据的新鲜度以及它们相对于应用程序上下文的准时性	19
Trustworthiness	可信度与源信誉和可靠性的概念相关联，并定义传感器反馈是否由真正的基础设施收集和处理	10
Availability	可用性是指传感器供电可运行和可用的时间	6
Security	安全性是保护传感器数据的隐私、机密性和完整性的尝试	5
Validity (freshness)	有效性或新鲜度是指在报告过程中，原始传感器中的读数还有多少是有效的	7
Frequency (temporal resolution)	频率(时间分辨率)显示关于特定时间的阅读速率	3

Confidence	置信度表明预期统计误差发生的最大概率	2
Volume (Throughput)	是指为达到目标信息而预期要处理的原始数据量	4
Ease of access	易访问性表明数据检索过程的难易程度	1
Interpretability	可解释性意味着数据具有有意义且易于解释的模式	1
Capacity	容量与并发访问有关，并显示并发的最大程度	1
Format	格式指的是“用户对信息呈现效果的感知”	1

上述表格内的评价维度均来源于各类文献对数据质量评价定义的不同概念，其中仍有重复之处，例如 **Accuracy** 和 **Precision** 同属于精度范畴，但在各自文献中的表述不同；此外，也有将更偏向于数据质量问题的维度混为数据质量评估使用，例如 **Volume** 等概念。

由于有关数据质量的评价维度目前没有统一的、较为信服的标准，对于每种指标，相应的文献里都能找到依据，其中，评价维度尚有相似之处，但维度的量化计算方法多种多样，没有统一的、相似程度较高的计算方式，文献^[1]也只能汇总出上述评价维度，同时文献^[1]对上述繁多的评价维度归类为评估数据质量的五个类别，这一点将在下一小节阐述。

2.2 数据质量问题

上文提到，文献^[1]的独特之处在于对数据质量的评估维度与数据质量问题做了区分，并指出这种区分是有必要的——以往的很多文章都忽略了这一点，很多文章将评估维度与数据质量问题混为一谈，或是将数据质量问题错误地作为评价指标使用。文中也对数据质量问题做了总结，如以下表 2.2 所示：

表 2.2 数据质量问题汇总

问题	描述	出现次数
Duplicate	冗余问题	4
Noisy data	-	2
Errors	主要指数据的语义出现错误	4
Lake of Context	缺乏上下文导致的语义模糊	2
Outliers	源于数据提供者和使用者的不匹配，当数据值显示与给定数据点的预期范围的任何随机偏差时，可能发生异常误差	3
Response time	数据得到完全处理的平均时间	1
Inconsistency	数据不一致问题	7
Redundancy	冗余问题	4
Uncertainty	数据离散缺失	3

上述表格中，“Errors”错误问题主要指数据的语义错误，例如在火灾温度传感器工作过程中，由于各种原因导致传感器周围温度过高而报警，但实际并无火灾发生，该误报问题即数据质量问题，即数据本身正确，但语义错误的情况。

“Inconsistency”数据不一致问题指相似环境下的多个传感器数据不一致问题，也是在保证数据质量的前提下如何处理数据的问题。不难发现，上述表格列举的数据质量问题大部分与数据的分析处理有关，以上问题也正是大部分有关解决方法的文献所要解决的经典问题，例如：去冗余问题、离散数据缺失问题、异常误报问题等。由此可以看出，数据质量的评估维度与数据质量问题确实是不同的，区分二者对于整个问题的研究十分有意义。

此外，文献^[1]还对上文总结的数据质量评估维度和数据质量问题做了横向的分类总结，将其划分为精度（Accuracy）、有效性（Confidence）、信誉度（Trustworthiness）、实时性（Timeliness）和完整性（Completeness）五个类别，并建议后续有关数据质量的因素都由此展开。五个类别的分类情况如以下表格2.3所示：

表 2.3 数据质量评价维度与数据质量问题分类

类别	数据质量评价维度	数据质量问题
Accuracy	Objectivity, Precision	Noisy, Error, Outlier
Confidence	Relevance, Ease of understanding, Interpretability, Format, Granularity	Uncertainty, Ambiguity, Lack of context
Trustworthiness	Reputation, Privacy, Security, Integrity	Insecurity, Source ambiguity, Inconsistency
Timeliness	Validity, Currency	High response time
Completeness	Usability, Ease of access, Availability, Throughput, Capacity, Frequency	Duplicate, Redundancy, Incompleteness

2.3 数据质量问题的解决方法

针对数据质量问题的解决方法多而复杂，有大量的现有工作在不同层面去解决，大部分文献针对数据质量问题中的某一个做了针对性的解决方法，在实用性上可以更快地解决具体问题。由于机器学习算法的热门及新方法对已有问题的“蓝海”领域，因此近几年的大部分文献中或多或少地与机器学习的方法有关，由此，本文将针对数据质量问题的解决方法按照以机器学习算法为主导的和非机器学习算法为主导的两个类别分别阐述。需要指出的是，并不是所有涉及到机器学习的算法的工作都被归类到机器学习为主导的类别中去，关键在于在解决文章

主要问题的方法中，机器学习的思想是否占主导地位，其次在分类中，将简单、初级的机器学习算法——例如聚类算法等也属于前者类别。

2.3.1 以机器学习为主导的方法

机器学习算法在数据处理、预测方面的优势有目共睹，通过训练，可以针对上述数据质量问题，例如数据冗余、误报、数据不一致等问题做有针对性的鉴别，以上可以归类为机器学习算法在数据处理上的应用；同时，机器学习在语义理解、上下文预测方面也有较大优势，因此本小节按照上述分类，在本小节内容的前几部分主要介绍有关数据处理的解决策略，后半部分着重介绍利用机器学习预测的优势应用于数据质量改进的方法。

文献^[5]主要针对数据上传带宽的问题，从减少端到云数据传输的角度出发，讨论如何减少数据上载量的问题，可以视为解决数据质量问题中的冗余问题。其核心工作是在端到云之间部署中间件来代替或提高物联网硬件的计算能力，将物联网设备产生的数据通过文中提出的一种自适应数据融合算法在中间件上计算后上传到云端。其中，自适应融合算法是一种融合提取多种特征值的改进聚类算法，融合后的数据上传到云端以起到减少数据传输量、融合特征一定程度上保护隐私的目的。通过上述介绍不难发现，该方法的不确定性在于：第一，如何在损害整体决策准确性的情况下融合传感器数据？第二，是考虑分布式处理的可行性，因为物联网中许多网络和边缘设备的处理能力较低，这是该方法面临的两大不足之处。

文献^[6]同样是基于聚类算法对数据进行预处理，其解决的主要问题同样可以归类于数据质量问题中的异常处理问题，该方法的特点在于预先将数据集的属性分为必要属性和非必要属性，在后续的处理中将非必要属性剔除，只在必要属性中提取特征值，以降低冗余减少干扰。其中，采用 K-Means 均值聚类算法对数据自动分类以提取特征，最后应用序列最小优化 (SMO) 监督算法过滤掉异常数据。文中详细地描述了算法步骤，并由实验证明了算法的效率与可行性，实验证明可有效降低数据冗余，提高数据的有效性。该方法将 K-Means 与 SMO 结合，由于聚类算法的存在，很难做到对数据的实时处理，因此属于典型的针对静态数据的处理方法。

文献^[7]针对数据质量评估，给出了一种权值动态变化的针对数据质量的评估系统——作者命名其为 Valid.IoT，该方法的核心即这套动态参数的评估模型。该方法同样利用了监督机器学习分类技术对参数做动态调整。此外，文中还提出了一种基于监督学习技术的差值算法，用来对模型检测出质量较低的数据做一致性调整，文中指出，该差值模型同样基于静态数据的训练，具体而言其训练数据来源于当前已有的判定为高质量的数据，不难发现，相较于动态参数的评估模型，差值模型对已有数据的训练其灵活性不如前者。

文献^[8]、文献^[9]、文献^[10]分别将机器学习在预测方面的优势应用于对数据质量的评估、数据值的预测以及数据语义的推理，取得了较好的效果。

文献^[8]针对数据质量的存储利用问题做了相应工作，即文中指出的两个问题：如何评估异构数据集质量与如何保持高质量的数据。文中首先对数据质量的定义做了探讨，作者认为，应当从数据消费者的角度来评估数据质量，文中给出了具体指标——数据的短时间、频繁访问程度。根据上述理论，作者提出了基于机器学习的数据过滤算法——QoS prediction framework，该算法的具体步骤为首先从数据的使用频率触发设定阈值，将数据分为高质量数据和低质量数据两种；在提取特征值上，低质量数据拥有更高的压缩比，对特征值的影响程度更低，而后通过聚类降低冗余，利用神经网络上下文过滤异常等环节，最终在云端存储的是精简处理后的特征值。此外，在云端的数据分析上，文中还指出在特征值上进一步划分局部特征和全局特征，例如：同一集群中的用户或同一集群中的服务具有相似的潜在集群特征，这些特征是局部特征；历史服务记录中探索用户之间和服务之间交互的深层潜在特征，这些潜在特征是全局特征，利用局部特征和全局特征评估数据间的相似度。

文献^[9]通过机器学习的预测特性对物联网数据质量进行预测，进而分析异构的物联网数据。具体而言，依据训练好的模型为物联网数据评估打分，以此作为数据质量的评估依据。该方法也属于针对静态数据的模型训练，同一般的模型训练相比，其独特之处在于，在训练的过程中，针对数据序列结构，采用一种自适应大小的滑动窗口检测法（AMWR 算法），窗口的大小可根据数据的上下文动态调整，具体做法包含以下四个步骤：初始选择最小窗口，选择带径向基函数核的支持向量回归机作为回归算法、提出了一种基于时间序列分析的新的通用方法（隆布-斯卡格尔方法）来寻找最佳窗口大小、预测窗口范围的大小。最后通过实验证实了该方法的效率与有效性。

文献^[10]所做的工作似乎更为庞大，针对物联网的异构数据流，构建了一个在现实物联网环境中运行的语义推理系统，用于从异构数据流中获取有价值信息。该推理系统包括：单一推理器、分布式推理器、移动推理器以及它们的混合。其针对的负载是有限的 RFID 数据，采用成熟的业界技术最终将其部署于移动终端、物联网节点及若干服务器上，实验通过比较各种语义数据格式下不同推理方法的可扩展性和实时性，验证了面向物联网应用的语义技术的能力。此外，我们评估了不同的数据聚合策略，以集成分布式物联网数据进行推理。基本实验结构如下图所示：

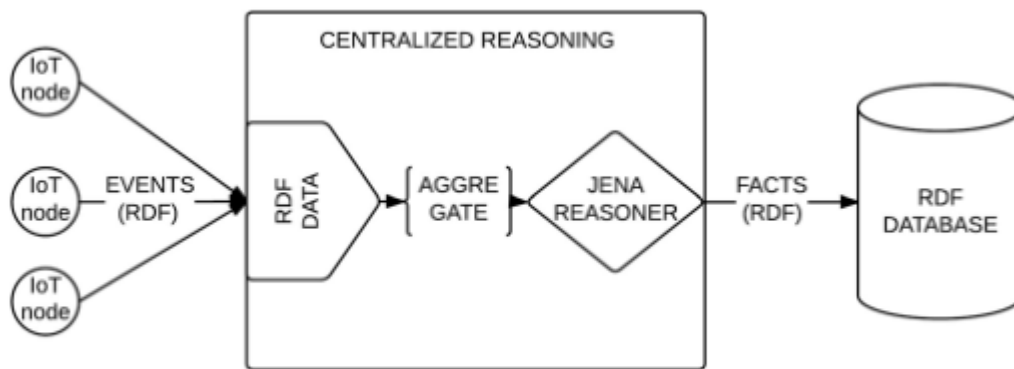


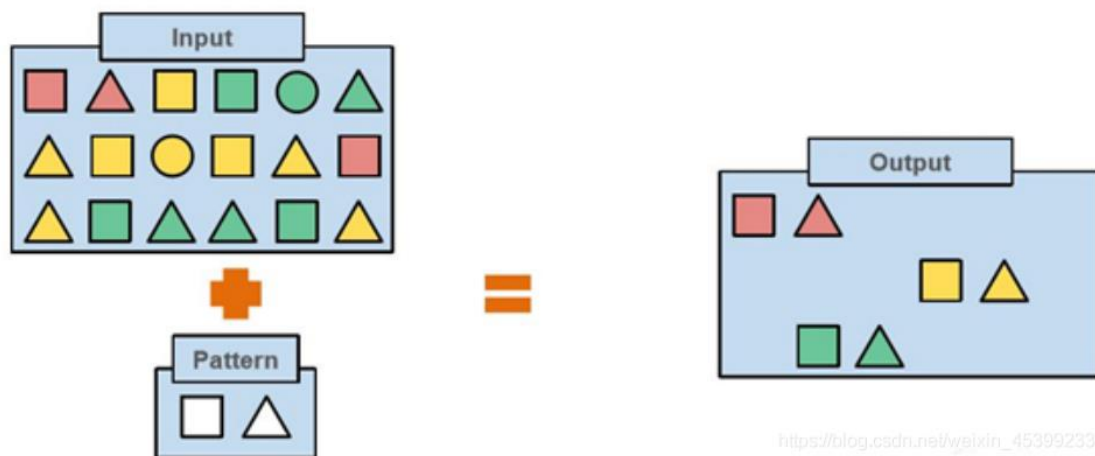
Fig. 1. Centralized reasoner.

2.3.2 以非机器学习算法为主导的方法

虽然近些年的文献偏重于机器学习算法的工作比较多,但同时也存在非机器学习相关的方法,比较而言,许多较为成熟的算法都没有应用机器学习,例如目前工业界已成熟应用的复杂事件处理系统(CEP),当然,CEP 严格上并不是真正的面向物联网的异构数据处理系统,而是在部署在云端中的应对不同结构数据流的事件发现系统,之所以将其提起,是因为其应用场景与物联网数据质量问题有相似之处,可以通过借鉴其思想为未来的研究提供思路。

本小节首先简要介绍 CEP。之所以称 CEP 为“应对不同结构数据流的事件发现系统”,是因为 CEP 的设计目的便是以设定好的事件模板为目标的模式匹配系统,如下图所示,其最大的特点便是实时性。CEP 从技术上严格说是一种针对传统数据库管理系统(DBMS)改进,更具体的说,是主要针对查询匹配功能上的改进,以应对异构数据流、实时动态查询的要求,因此,CEP 通常部署在云端。CEP 的实现原理可概括为以下几个步骤:设定匹配目标以供系统编译为目标事件、将异构数据流统一为一个运行时进行模式匹配、匹配到疑似目标后进入状态迁移、匹配到完整目标后触发事件。CEP 由于需要一定的计算资源,只能将其部署于云端,而其实时性、针对异构数据的分析性能使其成为一种能够有效应对物联网数据的实用系统。

CEP: Complex Event Processing



文献^[11]同样针对物联网数据的异常检测问题，不同的是，虽然该方法中包含机器学习的一些方法，但其主要思想是从概率的角度通过状态转移矩阵对单位窗口内的数据进行可信度计算，计算结果作为数据质量的评价指标。具体而言，文中提出了一种基于区域上下文的可信检测方法用于检测数据异常，文中首先针对异常提出了三种定义，分别为：点异常、行为异常和上下文异常。而后，基于区域上下文的可信检测方法具体步骤有：应用滑动窗口法对序列数据进行检测、为窗口内的数据可信程度进行计算、由阈值对计算结果进行过滤、最后利用机器学习的算法以过滤好的数据作为训练集训练概率检测器，最终得到检测模型。该方法的特点尤为明显，其优点是充分利用窗口内的上下文信息，其不足之处在于该方法有效的前提建立在以下假设之上——假设异常数据总是占总数据的一小部分，由此才能保证概率计算的有效性，同时，阈值在方法中起到过滤训练集的作用，其值的选取也很重要。

文献^[12]从工程角度说明了物联网数据面临的问题，同时该文章综述部分指出了现有的几项具有代表性的工作：FP7 COMPOSE（Collaborative Open Market to Place Objects at your Service）project 和 Ebbits project 等，前者的思路是设置开放市场，将终端抽象化为智能对象进而视为服务，可以使用标准的面向服务的计算方法对其进行管理，且可以动态组合，为最终用户提供增值应用；后者是一个基于开放协议和中间件的 SOA 平台，可以有效地将物联网子系统或设备转化为具有语义解析的 web 服务，方便企业将物联网集成到主流企业系统中。上述方法均在工程上有了良好实践。该文章的主要方法是对数据细分为元数据，并建立质量分析器和安全性分析器，前者主要从及时性、完整性、准确性和精确度等级打分构成向量，后者从机密性、完整性、隐私性、是否认证角度打分形成向量，同样地，两个打分机制采用学习机制的动态权值系统，最终为每项数据发布通道提

供一个分数。同时该文章还指出了数据流动的三层模型：接口层、处理层与发布层，这与文献^[13]中的思想相近，进一步证明了该模型的可用性。

3 总结与展望

本文针对引言部分提出的三个概念：数据质量、数据质量问题及数据质量问题的解决方案相应地提出了三个问题，即：如何评估数据质量、数据质量带来哪些问题及数据质量问题的解决方法有哪些。围绕三个问题，在本文的第二章分别做出了解答。

针对数据质量问题，目前的研究还没有一个统一的定义，相比于定义，更重要且更具有实际意义的问题在于给定统一的评估维度及量化评估维度。很遗憾，当前研究对此问题没有形成统一的评价方式，而是根据具体应用场景分别设计质量评估指标。本文第二章主要从概念与实际指标因素两方面阐述了区分数据质量评价维度与数据质量问题的必要性，指出数据质量的评价维度可以是客观通用的，而数据质量问题可以是有针对性的，待解决的实际问题。此外，对文献中的数据质量评价维度与数据质量问题做了总结。

针对数据质量问题如何解决，主要从是否为机器学习算法的类别出发，将现有工作概括为两个类别。以机器学习为主导的方法主要侧重于数据的处理与分析，聚类等算法在这一领域得到了广泛应用，同时，应用机器学习的预测特性由筛选后的数据作为模型的训练集也已有了许多工作，可以看出，由于机器学习在当前应用领域尚且处于“蓝海”状态，未来将有更多的此类应用、改进算法产生，尤其是在语义分析方向，虽然目前的语义分析只是建议一些形式化的推理模型，还不能像自然语言处理、知识图谱那样对数据内容有深层次的了解，但未来定将会有更多的机器学习的新技术应用于此方向。非机器学习的算法相对而言比较成熟，有些都已经得到大规模的应用，这些算法通常效率较高且效果稳定，且不像机器学习算法那样依赖于数据集，该类算法同样也具有很大的改进空间。

参考文献

- [1] MANSOURI T, MOGHADAM M R S, MONSHIZADEH F, et al. IoT Data Quality Issues and Potential Solutions: A Literature Review [J]. 2021.
- [2] GHALLAB H, FAHMY H, NASR M. Detection outliers on internet of things using big data technology [J]. Egyptian Informatics Journal, 2020, 21(3): 131-8.
- [3] KARKOUCH A, MOUSANNIF H, AL MOATASSIME H, et al. Data quality in internet of things: A state-of-the-art survey [J]. Journal of Network and Computer Applications, 2016, 73: 57-81.
- [4] BERTI-ÉQUILLE L. Measuring and Modelling Data Quality for Quality-Awareness in Data Mining [M]. Quality Measures in Data Mining. Springer Berlin Heidelberg. 2007: 101-26.
- [5] ALTURKI B, REIFF-MARGANIEC S, PERERA C. A hybrid approach for data analytics for internet of things [Z]. Proceedings of the Seventh International Conference on the Internet of Things. 2017: 1-8.10.1145/3131542.3131558
- [6] HOSSEINI M, BORJENI H R S. A Hybrid Approach for Anomaly Detection in the Internet of Things [Z]. Proceedings of the international conference on smart cities and internet of things - SCIoT '18. 2018: 1-6.10.1145/3269961.3269975
- [7] KUEMPER D, IGGENA T, TOENJES R, et al. Valid.IoT [Z]. Proceedings of the 9th ACM Multimedia Systems Conference. 2018: 294-303.10.1145/3204949.3204972
- [8] GAO H, XU Y, YIN Y, et al. Context-Aware QoS Prediction With Neural Collaborative Filtering for Internet-of-Things Services [J]. IEEE Internet of Things Journal, 2020, 7(5): 4532-42.
- [9] AKBAR A, KHAN A, CARREZ F, et al. Predictive Analytics for Complex IoT Data Streams [J]. IEEE Internet of Things Journal, 2017, 4(5): 1571-82.
- [10] MAARALA A I, SU X, RIEKKI J. Semantic Reasoning for Context-Aware Internet of Things Applications [J]. IEEE Internet of Things Journal, 2017, 4(2): 461-73.
- [11] ZHAO B, LI X, LI J, et al. An Area-Context-Based Credibility Detection for Big Data in IoT [J]. Mobile Information Systems, 2020, 2020: 1-12.
- [12] SICARI S, RIZZARDI A, MIORANDI D, et al. A secure and quality-aware prototypical architecture for the Internet of Things [J]. Information Systems, 2016, 58: 43-55.
- [13] BALI A, AL-OSTA M, ABDELOUAHED G. An Ontology-Based Approach for IoT Data Processing Using Semantic Rules [M]. SDL 2017: Model-Driven Engineering for Future Internet. Springer International Publishing. 2017: 61-79.