This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2773600, IEEE Internet of Things Journal

Context Aware Computing, Learning and Big Data in Internet of Things: A Survey

Omer Berat Sezer, Erdogan Dogdu, Ahmet Murat Ozbayoglu

***Abstract*—Internet of Things (IoT) has been growing rapidly due to recent advancements in communications and sensor technologies. Meanwhile, with this revolutionary transformation, researchers, implementers, deployers, and users are faced with many challenges. IoT is a complicated, crowded, and complex field; there are various types of devices, protocols, communication channels, architectures, middleware, and more. Standardization efforts are plenty, and this chaos will continue for quite some time. What is clear, on the other hand, is that IoT deployments are increasing with accelerating speed, and this trend will not stop in the near future. As the field grows in numbers and heterogeneity, "intelligence" becomes a focal point in IoT. Since data now becomes "big data", understanding, learning, and reasoning with big data is paramount for the future success of IoT. One of the major problems in the path to intelligent IoT is understanding "context", or making sense of the environment, situation, or status using data from sensors, and then acting accordingly in autonomous ways. This is called "context aware computing", and it now requires both sensing and, increasingly, learning, as IoT systems get more data and better learning from this "big data". In this survey, we review the field, first, from a historical perspective, covering ubiquitous and pervasive computing, ambient intelligence, and wireless sensor networks, and then, move to context aware computing studies. Finally, we review learning and "big data" studies related to IoT. We also identify the open issues and provide an insight for future study areas for IoT researchers.**

Keywords*—Internet of things, Context Awareness, Machine Learning in IoT, Big Data in IoT, Data Management and Analytics .

I. INTRODUCTION

Internet of Things (IoT) is the umbrella phrase covering all sorts of things connected to internet. These "things" include everything from dummy sensors, like motion sensors, temperature measuring devices, etc., to various types of smart things such as smart phones, smart meters, autonomous cars, buildings, etc. The idea is that all these "things" will collect data, share data and information, and at the end everything in this ecosystem (systems, people, etc.) act accordingly in smart ways so that our lives are easier, better and in harmony.

The IoT paradigm started with Radio Frequency Identification (RFID) and sensor network technologies. In 1999, this term was first mentioned by Kevin Ashton [1]. However, the definition has been changing with the evolving technology, and lately, it has become "The Internet of Things allows people and things to be connected Anytime, Anyplace, with Anything and Anyone, ideally using Any path/network and Any service" [2]. In other words, intelligent devices (things) sense their surrounding environment, understand changes, interact with each other, and analyze the results by using existing Internet infrastructure and standards. Today, IoT is widely used in different areas such as transportation, healthcare, and utilities.

The projected impact of IoT devices usage is tremendous. The US National Intelligence Council (NIC) predicts that "by 2025 Internet nodes may reside in things that we use everyday, food packages, furniture, paper documents, and more" [3]. In the future, almost all active devices will have an Internet interface. This vision enforces data scientists to provide solutions for the inevitable challenge of how to process the large amount of data coming from things, and how to make sense of the raw extracted data. The IoT paradigm covers a vast amount of different areas and poised to attack various existing and upcoming problems, such as hardware, power, security, reliability, interoperability, and data sharing problems. Several different architectures and middleware solutions are proposed for attempting to solve these aforementioned problems. In addition, there are also other study areas like context awareness, semantic and cloud computing, reasoning and processing, and data and service management, associated with the IoT concept.

To better understand the studies and technologies regarding the IoT, it might also be beneficial to examine related areas, such as ubiquitous computing, pervasive computing, ambient intelligence (AmI), smart homes and cities, machine to machine (M2M) communication, wireless sensor networks (WSNs), semantic sensor networks (SSNs), web of things (WoT), context awareness, semantics and big data, machine learning, and data mining.

Ubiquitous computing, pervasive computing, and AmI were proposed before the era of IoT studies and technologies. In the late 1980s, researchers studied the human-to-human interface using technology, and as a result, ubiquitous computing was formed. Mark Weiser, the father of the idea, defined ubiquitous computing and the smart environment as "the physical world that is richly and invisibly interwoven with sensors, actuators, displays, and computational elements, embedded seamlessly in the everyday objects of our lives, and connected through a continuous network" [4][5]. This vision supporting the interconnection of embedded devices and computers was a pioneer in the development of the Internet. Moreover, the ubiquitous computing idea provided an inspiration showing that computing is not limited only to one platform, but is embedded with

Omer Berat Sezer is a Ph.D. Candidate at Department of Computer Engineering, TOBB University of Economics and Technology, Ankara, 06560 Turkey e-mail:oberatsezer@etu.edu.tr

Prof. Dr. Erdogan Dogdu is with the Department of Computer Engineering, Cankaya University, Ankara, 06790 Turkey e-mail:edogdu@cankaya.edu.tr

Assist. Prof. Dr. Murat Ozbayoglu is with the Department of Computer Engineering, TOBB University of Economics and Technology, Ankara, 06560 Turkey e-mail:mozbayoglu@etu.edu.tr

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2773600, IEEE Internet of Things Journal

any device, anywhere, and anytime. These are the key terms of the IoT.

Pervasive computing is a term that is defined as "Connected computers and sensors are communicating with each other to understand the surrounding environment". In the late 1990s, researchers proposed another term called "ambient intelligence" that was expressed as "electronic devices support people for everyday activities by sensing and responding to them according to the environmental changes". AmI most frequently focuses on home consumer electronics and their implementations. In the 2000s, companies began to focus on smart home and home automation. Their applications and products were mostly based on the communication of home appliances, and responding to people via wired and wireless technology. Smart home technologies generally focus on home appliances. On the other hand, M2M technologies propose more comprehensive solutions in such a way that the machines can communicate with each other directly using wired or wireless technologies. These solutions and terms form the basis of the IoT.

In the 2000s, the researchers began to study WSN, which is considered to be the ancestor of the IoT. In a WSN, different types of sensors, such as temperature, pressure, sound, each behaving as a unique node, create networks together to pass data through the network. WSNs can consist of hundreds or thousands of nodes. The size of a WSN depends on the application area and purpose of the application. SSN is a particular type of WSN that combines the semantic web and sensor network. The data and descriptions from the sensors are encoded, defined, and expressed comprehensively via semantic web languages. WSN and SSN are the subset areas of IoT. In "Related Areas" section, all related terms and studies on these subjects will be explained in detail.

In this paper, we also cover "the extended IoT" research areas, namely Web of Things (WoT) and Semantic WoT (SWoT). WoT extends IoT with standard Web protocols such as the standard URL access to things, HTTP communication protocol with things, and other standard Web protocols such as JSON for data format. This way IoT will be more integrated with the ubiquitous Web. In 2007, several researchers (Dominique Guinard and Vlad Trifa are among them) proposed the WoT and published the first manifesto of the WoT [6] [7]. SWoT on the other hand further extends WoT by using standard Semantic Web protocols, so that data obtained in IoT is semantically meaningful and more interoperable within the ecosystem. Things are encoded through semantic web languages (RDF, OWL, etc.) to handle the interoperability problem between ontologies and data [8] [9]. The overall goal of these efforts is to make all IoT components more interoperable.

Some of the IoT research areas have overlaps with other fields such as context awareness, big data analytics, machine learning, and data mining. Context awareness is a term that is used for representing the case where computer and embedded devices sense and react according to the changes in their environment. First introduced by Schilit in 1994 [10], a context aware system acquires, understands, recognizes the context, and takes an action according to that particular context. For the IoT side, Perera et al. [2] completed a survey on context-aware computing in IoT with a comparison of 50 different projects. In the "Context Awareness" section, all related terms and studies on this subject will be covered in detail.

Big data is another overlapping study area where the corresponding methodologies for processing large data sets, which cannot be processed through traditional data processing techniques, are examined. In real life, IoT data sets can easily have large volumes, varieties, and velocities due to the nature of data. In addition, IoT has the following features: "intermittent sensing, regular data collection, and sense-compute-actuate (SCA) loops" [11]. Hence, IoT converges to Big Data to analyze data and make inferences from collected datasets. According to a paper published by Zaslavsky et al., it is expected that the total amount of data on earth will reach up to 35 zettabytes (ZB) in 2020 [11]. In "Big Data" section, all related studies on this subject will be discussed in detail.

Besides these topics, machine learning and data mining also have some common studies that overlap with IoT. In 1950s, Arthur Samuel defined machine learning as a "field of study that gives computers the ability to learn without being explicitly programmed" [12]. While analyzing the IoT data, in order to get better results and higher overall performance, it might be necessary to include some predictive analytics within the system. Machine learning algorithms have been studied extensively in the last few decades and research shows that these algorithms provide better predictions and decisions with more available data collected from different sources. In addition to machine learning, data mining also supports better predictions and decisions through using appropriate learning algorithms. In the "Machine Learning" and "Data Mining" sections, all related terms and studies on these subjects in conjunction with the IoT will be explained in detail.

This survey paper focuses on the IoT related subjects: "context awareness", "inferences from context", "context reasoning", "learning algorithms" to make predictions, profiling, and data analysis using big data. When the literature is reviewed, IoT related survey papers can be grouped into the following categories:

- IoT general purpose surveys and open issues [4] [13] [14] [15] [3] [16] [17] [18]
- Survey of IoT platforms and frameworks [19] [20] [21] [22],
- Survey of context awareness [23] [24] [25] [21] [26] [27] [28] [29],
- Survey of machine learning on specific topics (Human Activity Recognition, Mobile Phones Sensing, Body Sensor Networks, Wireless Sensor Network) [24] [30] [31],
- Survey of data mining [31] [32] [27],
- Survey of IoT related areas (Sensor Network, Social IoT, Mobile Phone Sensing) [33] [34] [35] [36],
- Survey of big data analytics [37] [38].

This survey article covers the IoT related literature from both historical and conceptual perspectives for context awareness, machine learning, and big data. None of the existing survey papers covers these closely-related fields all together from a fully-functional framework or system of systems perspective. However, recent advancements in sensors, IoT technologies,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2773600, IEEE Internet of Things Journal

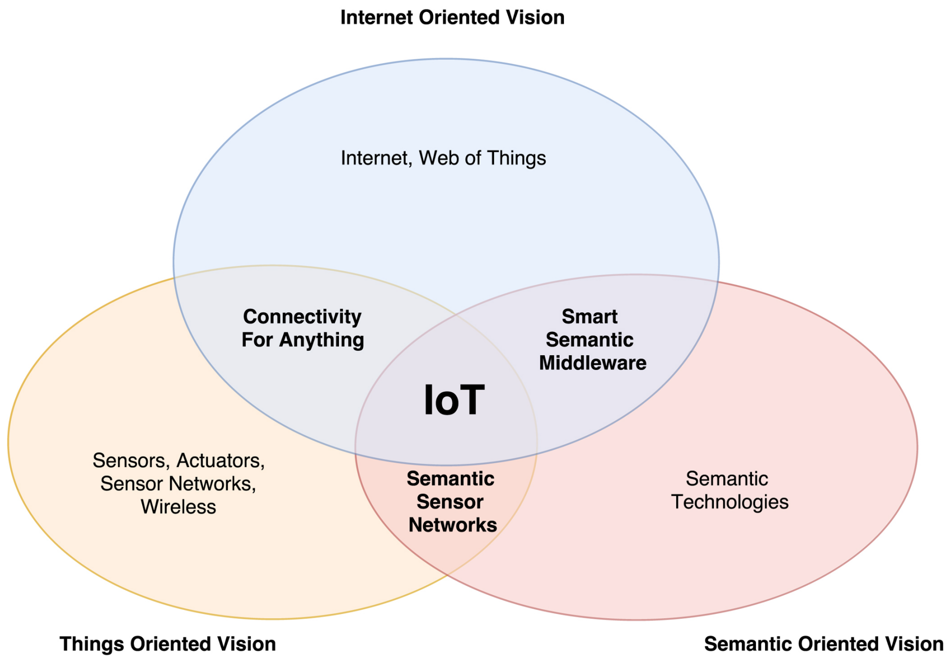

Fig. 1. Different Visions Intersection: Internet of Things, Adapted From [3]

and big data systems created the opportunities for the emergence of machine and deep learning-based, real-time predictive, and preventive maintenance and analytics systems. To the best of our knowledge, no single survey paper has covered these topics together even though they are highly related. In addition, this survey paper consists of a comparison of IoT platforms and frameworks. Furthermore, IoT machine learning and big data topics are also examined and surveyed together because IoT big data and streaming data are analyzed using machine learning methods.We also review new and recent advancements in IoT. Open Issues and future directions for IoT machine learning and IoT big data are also discussed.

The rest of this paper is organized as follows: In Section II, we introduce IoT under the following headlines: IoT paradigm and definition, IoT potential and application domains, IoT characteristics and features, and IoT research trends, areas, and open issues. In Section III, related areas and history of IoT technologies, in regard to ubiquitous computing, pervasive computing, AmI, WSNs, WoT, and social internet of things, are presented. In Section IV, context awareness is analyzed in the following subsections: Context and context awareness; Context aware features and context types; and Context life cycle in IoT. In Section V, machine learning algorithms for IoT are categorized as supervised, unsupervised, and reinforcement learning. In addition to these topics, proposals and studies in the literature are compared and analyzed. In Section VI, IoT data processing is evaluated through the vision of big data. In Section VII, open issues in context awareness, machine learning and big data analytics in IoT are evaluated. We then conclude in Section VIII.

II. Internet of Things (IoT)

A. IoT Paradigm and Definition

The IoT is a network of internet connected physical objects - embedded devices, vehicles, sensors, computers; and these objects exchange data among themselves and other systems [39]. Another definition of IoT is “The Internet of Things allows people and things to be connected Anytime, Anyplace, with Anything and Anyone, ideally using Any path/network and Any service” [2]. In fact, IoT stands for “a world-wide network of interconnected objects uniquely addressable, based on standard communication protocols” [40]. The number of such addressable objects is growing rapidly.

The IoT paradigm is also defined as an intersection of three vision areas: Internet oriented vision, things oriented vision, and semantic oriented vision [3]. Figure 1, which is adapted from [3], shows these vision areas and their intersections. Sensors, actuators, sensor networks, RFID, near field communications (NFC), electronic product code (EPC) technologies, wireless sensor, and actuator networks (WSAN) are classified as part of the things oriented vision. Semantic technologies (web, languages, execution environment, and so on), reasoning over data, are classified as a semantic oriented vision. Finally, Internet technologies and the WoT are located in Internet oriented vision. There are also some technologies in the overlapped regions of these visions. For example, semantic sensor networks are in the intersection area of the things and semantic oriented visions. The intersection of all three vision areas is the IoT.

B. IoT Potential Application Domains

The IoT has a huge market potential with an increasing growth rate in recent years. In 2014, it was estimated that there were 1.5 billion Internet-enabled computers and 1 billion Internet-enabled mobile phones. By 2020, the number of Internet connected IoT devices will be somewhere between 50 to 100 billion [2] and the total amount of data generated by humans and devices on earth will reach 35 zettabytes (ZB) [11] [41]. According to BBC Research, the global market for sensors was $56.3 billion in 2010 [11], and reached $101.9 billion in 2015. It is expected that the market will reach $113.2 billion by 2016 and $190.6 billion by 2021 [42]. According to the Gartner Research report, the number of IoT devices will reach 25 billion in a few years [27] [43].The IoT is used in many different application domains. An alphabetically sorted short list follows below [44] [3] [45] [27]:

- Aerospace and aviation (manufacturing: detection of suspected unapproved parts),
- Education (remote education, enhanced reality for learning),
- Energy (smart metering, coordination of generation and storage),
- Entertainment and sports (gaming, sports, cinema, smart gym),
- Environment (chemical detection, temperature, and humidity monitoring),
- Finance and banking (POS terminals, remotely located ATMs, on-line desktop, and mobile device banking),
- Food and farming (controlling and monitoring facilities, monitoring of produce, livestock, defect management, protecting chemical and environmental conditions, automation of ordering service, automation of delivery process and accounting),
- Government (real-time environmental monitoring, remote service delivery, asset tracking, smart cities, city and building management and security),
- Healthcare (remote treatment and surgery, remote diagnostics and examinations, remote patient monitoring and tracking, medical asset tracking),
- Home automation (smart appliances, home security, and monitoring),
- Information and communication technologies (security and monitoring, remote management, device tracking, and automation),
- Logistics (mobile ticketing),
- Manufacturing and heavy industry (process monitoring and management, equipment monitoring, shipping tracking, remote servicing, monitoring employees, suppliers, inventory management),
- Pharmaceutical industry (drugs tracking, pharmaceutical products, security and safety),
- Public safety and military (surveillance network, remote asset control and tracking, disaster management),
- Retail and hospitality (anti-theft and fraud, facilities monitoring and management),
- Transportation (smart roads, rails, runways, assisted driving, traffic signals, augmented maps, intelligent transportation systems),
- Vehicles (smart bus, planes, boats, trains, automobiles),
- Water (system pressure, reservoir levels, water quality).

C. IoT Characteristics and Features

The IoT has different characteristics when compared to sensor networks. According to [21], there are seven main characteristics: architecture, complex system, everything as a service, intelligence, size considerations, space considerations, and time considerations.

Although there are different architectures for the IoT, two main architectures are more relevant: event-based and time-based. Event-based architectures are triggered with events (e.g., motion detection, window open). On the other hand, time-based architectures produce data continuously (e.g., humidity and pressure sensors) [55].

In the IoT world, there are billions of devices. Some of them have high-level architectures with large memory capacities, high CPU speed, and high reasoning capabilities (e.g., mobile phones). Meanwhile, some of them, on the contrary, have low-level architectures, limited memory, and computing capabilities (e.g., temperature sensors). The interconnections between these devices make the IoT a very complex system, in general.

Cloud computing is expected to play an important role in the IoT ecosystem. With cloud computing, IoT storage and computing capacities can be increased in a scalable manner. Moreover, sensors can be used everywhere and processing sensor data can be accomplished through cloud computing services. Everything as a service, proposed by [56] is a cloud computing term that all services (Main services: infrastructure as service, software as service, platform as service. Others: desktop-as-service, storage-as-service, database-as-service) are collected in one hand. These services fulfill the needs of the IoT, including infrastructure, sensing-as-service [11] and more.

In the IoT infrastructure, data knowledge extraction can be implemented by collecting, modeling, and reasoning over the massive amounts of data collected from devices. This is simply described as the intelligence of IoT. Context awareness is further added to the picture with the fusion of sensor data, modeling, and reasoning about context. It is predicted that the number of Internet connected IoT devices will be 50 to 100 billion by 2020 [2]. The interaction of billions of interconnected devices will therefore cause problems leading to size considerations for IoT environments and big data problems that can be solved with cloud computing, as mentioned above.

Another important characteristic of IoT is space consideration, including time and location. While extracting and evaluating context awareness information from sensor data, the location, time, and duration of data become important factors for IoT processing. The number of IoT devices is increasing rapidly in time, and tracking these devices will be more and more difficult [2].

According to [14], there are also system-level features that IoT should support. Heterogeneity of devices, scalability, data exchange through wireless technologies, energy saving, tracking capabilities, self-organization capabilities, security and privacy-preserving, semantic interoperability, and data management are some of the noteworthy features.

TABLE I. IoT RESEARCH AREAS, BRIEF INFORMATION AND RECOMMENDATION

IoT Research Areas	Brief Information - Recommendation
Addressing and Naming Devices	Reaching specific devices efficiently in IoT world is an issue to be solved [3] [16]. Social IoT Networks are proposed [46]
Big Data	Big Data storage, streaming, processing and analytics are the subtopics of this issue. Related research topics are data mining, knowledge discovery, data quality and uncertainty, transaction handling, semantic event processing and semantic enrichment [32].
Context Aware Computing	Context aware computing is about sensing the environment and context, and adapting the behaviour accordingly in IoT systems [21].
Data Management	Management of large, distributed, and heteregenous IoT data is challenging task [16] [22].
Device Management	Remotely configuring, monitoring, updating and controlling IoT devices are the subtopics of this issue. New frameworks and solutions are developed.
Energy Management	Energy usage and battery lifetimes are the main problem of the resource constrained IoT devices. There are research studies on the development of enery optimized solutions [14] [36] [22].
Hardware Development	Development of IoT devices, sensors, RFID tags is an important research area that researchers and engineers study. There are new researches and studies on IoT devices and scenarios such as smart home, health care, biomedical sensing, environmental monitoring, tranportation [47] [48] [49] [50] [51] [52].
Information Centric Network (ICN)	ICN is a new approach for Internet architecture. It uses content for addressing data instead of using Ip address. There are new researches and studies on ICN-IoT application and scenarios [53] [54].
Interoperability and Heterogeneity	Many different kinds of IoT devices are produced and developed in recent years. Communication of these devices cause interoperability problems. Semantic solutions are trying to resolve this issue [14] [36].
M2M Communication	Routing, end to end reliability, development of new network protocols are the research areas in the communication of the IoT devices [16].
Machine Learning and Artificial Intelligence	Machine Learning is required to design self organized, managed, adapted, processed IoT devices [16] [36] [22] [26].
Resource Management	In IoT world, there is limited access to resources. Resource management is needed to design and develop efficient IoT systems.
Security and Privacy	Security and privacy are central and important topics in IoT systems. Secure authorization and authentication, security of IoT data, secure access are the sub research areas of these issues [16] [14] [3].
Service Management	IoT devices are distributed and located everywhere. Service Oriented Architecture (SOA) is required to manage and discover devices.
Standardization Activities	New IoT standarts and protocols are still emerging and continue to be developed [14] [3] [22] [26].
System and Network Architecture	Researchers are trying to develop scalable, efficient, robust, cost-efficient IoT system and network architectures [16] [14].

D. IoT Standards

There are many IoT related standardization activities in literature. A number of standards organizations such as IEEE, International Organization for Standardization, International Telecommunication Union, International Electrotechnical Commission, European Committee for Electrotechnical Standardization, China Electronics Standardization Institute, and American National Standards Institute are working on the IoT standardization activities [13]. For example, IEEE lists a long list of their IoT related standards in their web site[1]. Standardization activities in this area can be classified as RF layer, lower layer, communication layer, data protocols layer, semantic and higher layer standardization activities. IoT technologies mostly use these standards.

RF layer and near field communication interface and protocol (NFCIP) are standardized (ISO 18092, 21481, 22536 and 23917; ECMA 340, 352, 356 and 365; ETSI TS 102 190) by various authorities, such as the International Organization for Standardization (ISO), European Computer Manufacturers Association (ECMA), Global System for Mobile Communications Association (GMSA), and European Telecommunications Standards Institute (ETSI) [14].

Lower layer (PHY, MAC) is standardized by IEEE (IEEE 802.16 Wireless Broadband Standards, IEEE 1547 Standard for Interconnecting Distributed Resources with Electric Power Systems, IEEE 1609 IEEE Wireless Access in Vehicular Environments (WAVE), IPv6 over low-power WPAN (6LoW-PAN); the routing over low-power and lossy networks (ROLL), Ethernet) by The Institute of Electrical and Electronics Engineers Standards Association (IEEE-SA) and Internet Engineering Task Force (IETF) [14].

Communication layer is also standardized by IEEE (IEEE 802.15 Wireless Personal Area Network, IEEE 802.15.1 WPAN / Bluetooth, IEEE 802.15.4 Low-rate wireless personal area networks, IEEE 802.15.6 Wireless Body Area Network). Zigbee (standardized with IEEE 802.15, IEEE 802.15.4). Other IoT related standards and technologies are listed as follows: Zwave (Standard: Z-Wave Alliance ZAD12837 / ITU-T G.9959), Insteon, NFC, Narrow-Band IoT (NB-IoT), LTE-Machine Type Communication (LTE-MTC).

Data Protocol layer is also standardized by a number of organizations. Data protocol standards related to IoT technologies can be listed as follows: Message Queuing Telemetry Transport (MQTT)[2], MQTT For Sensor Networks (MQTT-SSN), Constrained Application Protocol (CoAP)[3], The Simple Text Oriented Messaging Protocol (STOMP)[4], Exten-

[1] http://standards.ieee.org/innovate/iot/stds.html

[2] http://mqtt.org/

[3] https://datatracker.ietf.org/doc/rfc7252/

[4] https://stomp.github.io/implementations.html

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2773600, IEEE Internet of Things Journal

sible Messaging and Presence Protocol (XMPP)[5], XMPP-IoT, Advanced Message Queuing Protocol (AMQP)[6], Data-Distribution Service for Real-Time Systems (DDS), Java Message Service (JMS), Lightweight local automation protocol (LLAP), Lightweight M2M (LWM2M), Simple Sensor Interface (SSI), Representational State Transfer (REST), Simple Object Access Protocol (SOAP), Websocket.

IoT related technologies also use semantic standards such as IOTDB[7], SensorML, Semantic Sensor Net Ontology, RESTful API Modeling Language(RAML)[8], Media Types for Sensor Markup Language(SENML), Lemonbeat smart Device Language (LsDL).

Higher level IoT layer is standardized by ETSI, Object Management Group (OMG), AllSeen Alliance, World Wide Web Consortium (W3C), Semantic Sensor Network Incubator Group, OneM2M Partners, Industrial Internet Consortium, Open Interconnect Consortium - Open Connectivity Foundation. OneM2M[9] is a high level standardization activity on IoT devices, working on an interoperability framework towards a common M2M or IoT Service Layer for all types of devices and framework. OneM2M organization consists of eight leading standards bodies from different countries and over 200 members. AllSeen Alliance is developing an open source software framework, called AllJoyn[10], that provides IoT device communication and management functions. In addition to these, there are also high level frameworks and solutions such as IoTivity, IEEE P2413, Thread, IPSO Application Framework, OMA LightweightM2M v1.0, and Weave.

E. IoT Research Trends and Areas

IoT is in the intersection of many different visions and disciplines. Thus, it contains several subtopics and research areas. Many articles in the literature classify and study these research areas and trends in many different subtopics. We list the major IoT related research areas in Table I with a brief explanation and relevant references.

According to [3], research topics and open issues can be classified into standardization activities, addressing and networking issues, and security and privacy. In the literature, many different standards are mentioned, such as the Internet Engineering Task Force (IETF), European Telecommunications Standards Institute (ETSI) standards, and others. However, there is no single integrated framework to combine all different standards. Therefore, there are several devices working with different standards and frameworks within their domain, and this leads to interoperability problems between heterogeneous systems. This problem could be solved with a joint and unified set of standards and frameworks.

According to [15], research areas for IoT are categorized as massive scaling, architecture and dependencies, creating knowledge and big data, robustness, openness, security, privacy, and humans in the loop. The number of internet connected devices is increasing rapidly, hence current existing solutions, protocols, and research, in terms of meeting expectations and requirements need to be re-evaluated continuously. Massive scaling, architecture, and dependencies focus on these problems. With ever increasing devices, collected data also increase and create big data that is not immediately semantically meaningful. Meanwhile, it is expected that knowledge can be extracted from this massive data. These problems can be solved with data mining and machine learning algorithms in scalable architectures. Another research area is openness. Traditionally, all devices and vehicles, which have sensor-based systems, have closed loop systems, and that is no longer convenient for many real-world IoT scenarios. These systems should be open for interoperability to solve the security problems as well.

Authors of [13] collected research trends under these subheadings: Integrating IoT solutions to social networking, developing green technologies, developing context aware IoT middleware solutions, using artificial intelligence techniques to create intelligent things, combining cloud computing and IoT. Different research areas and paradigms emerged over the years. One of them was integrating social networking with IoT solutions. Atzori et al. [46] proposed the SIoT paradigm that connects IoT devices with each other using social networks. Other paradigm is the WoT [57] [58] [8] that connects devices by using existing and well-known web standards. SIoT and WoT will be discussed in detail in Section III. Another issue is the reduction of power consumption for IoT devices and sensors. Energy efficient techniques should be developed to save energy [59]. Billions of IoT sensors and devices connect to the Internet and other systems and require semantically meaningful and context awareness capabilities. Most IoT middleware solutions do not have context aware capability [13]. The European Union hopes that context aware solutions will be proposed in the time frame of 2015-2020 [60]. Some of the context aware studies in IoT are summarized in [27]. Another research area is the addition of artificial intelligence to IoT devices and systems. Bringing artificial intelligence to the IoT was proposed by Arsnio et al [61]. Futuristic ideas for IoT are also emerging and defined as self-configurable, self-optimized, self-protected, and self-healed IoT devices and systems [62] [63]. Combining cloud computing and the IoT accelerate IoT computing capabilities. Sensing as a cloud service structure is an example solution of combining cloud computing and the IoT [4][64].

As previously mentioned, there are many research areas and open issues, however, we focus on context awareness, drawing inferences from context, context reasoning, and learning algorithms to make predictions and profiling, and also data analysis using big data. The main research areas, studies, and issues will be reviewed in Section IV for context awareness, Section VI for big data, and Section V for machine learning. Open issues and possible future directions for research are discussed in Section VII.

[5] https://xmpp.org/
[6] http://www.amqp.org/
[7] https://iotdb.org/
[8] http://raml.org/
[9] http://www.onem2m.org/
[10] https://allseenalliance.org/framework

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2773600, IEEE Internet of Things Journal

TABLE II. IoT PLATFORMS (DEVICE M: DEVICE MANAGEMENT, DATA M: DATA MANAGEMENT, RT ANALYTICS: REAL TIME ANALYTICS, BD ANALYTICS: BIG DATA ANALYTICS, LT: LEARNING TOOL)

Platforms	Web Sites	Device M.	Data M.	RT Analytics	BD Analytics	LT
AllJoyn	https://allseenalliance.org/framework	✓	✓			
AirVantage	https://airvantage.net/	✓	✓		✓	
Arkessa	http://www.arkessa.com/	✓	✓	✓		
ARMmbed	https://www.mbed.com/	✓				
Brillo	https://developers.google.com/brillo/	✓	✓			
Carriots	https://www.carriots.com/	✓	✓	✓		
Devicehub.net	https://www.devicehub.net/	✓	✓			
Everyware Device Cloud	http://www.eurotech.com/	✓	✓	✓		
EvryThng	https://evrythng.com/	✓	✓	✓		
Exosite	https://exosite.com/	✓	✓	✓		
GroveStreams	https://grovestreams.com/	✓	✓	✓		
Ericsson IoT-Framework	https://github.com/EricssonResearch/iot-framework-engine		✓	✓		
IFTTT	https://ifttt.com/		✓			
IoTivity	https://www.iotivity.org/	✓	✓			
Intel IoT Platforms	https://software.intel.com/	✓	✓	✓		
LinkSmart	https://linksmart.eu/redmine		✓			
NinjaBlock	https://ninjablocks.com/	✓				
OpenIoT	http://www.openiot.eu/		✓			
OpenMTC	http://www.open-mtc.org/	✓	✓	✓		
Open.Sen.se	http://open.sen.se/		✓	✓		
Pentaho	http://www.pentaho.com/internet-of-things-analytics		✓	✓	✓	
realTime.io	https://www.realtime.io/	✓	✓	✓		
SensorCloud	http://www.sensorcloud.com/		✓	✓		
SkySpark	http://skyfoundry.com/skyspark/		✓	✓		
Statistica	http://software.dell.com/products/statistica/		✓	✓	✓	
Tellient	http://tellient.com/index.html		✓	✓		
TempoIQ	https://www.tempoiq.com/		✓	✓		
The thing system	http://thethingsystem.com/	✓				
ThingSpeak	https://thingspeak.com/	✓	✓	✓		
ThingSquare	http://www.thingsquare.com/	✓	✓	✓		
ThingWorx	https://www.thingworx.com/	✓	✓	✓		
Sense Tecnic WoTkit	http://sensetecnic.com/		✓	✓		
Watson IoT Platform	http://www.ibm.com/internet-of-things/iot-solutions/watson-iot-platform/	✓	✓	✓	✓	✓
Xively	https://xively.com/	✓	✓			
Vitria	http://www.vitria.com/iot-analytics		✓	✓		
Weave	https://developers.google.com/weave/	✓				

F. IoT Platforms, Frameworks, Services and Middleware

There are many IoT platforms, frameworks, services, and middleware that can collect, process, and analyze sensor data. Our aim is not to review and survey existing platforms and frameworks. Articles [65] [20] [19] surveyed existing IoT platforms, frameworks, systems, prototypes, middleware, and different approaches. Some of the important IoT platforms are listed in Table II. They are listed with their device management, data management, real-time analytics, big data analytics, and learning capabilities. It is clear that many existing platforms have limited analytics and learning capabilities.

G. IoT Security and Privacy

IoT networks all over the world consist of billions of devices. Security and privacy issues are raised due to the large number of devices and lack of unified standardization studies on IoT security. Every internet connected device can easily cause security problems. Hence, these studies should not be ignored. In recent years, Denial-of-service (DoS) attacks have shown that research and implementation of IoT security solutions are very important. In this paper, we do not aim to review and survey all IoT security issues and research areas.

In literature, there are many papers covering IoT security issues [66][67][68][69]. Abomhara et al. [66] surveyed different types of threats related to IoT in terms of intruder model, denial-of-service attacks, physical attacks, eavesdropping and passive monitoring, traffic analysis, and data mining. They also examined the security and privacy challenges in IoT with the following subheadings: user privacy and data protection, authentication and identity management, trust management and policy integration, authorization and access control, end to end security, attack resistant security solutions. In addition, Mahmoud et al. [67] defined security principles that should be enforced to achieve security between IoT devices and people in regards to confidentiality, integrity, availability, authentication, lightweight solutions, heterogeneity, policies, and key management systems. They also defined the security challenges in each layer of IoT devices and IoT security countermeasures in their survey paper. Zhang et al. [68] summarize ongoing research studies in IoT security with the following topics: object identification, authentication, authorization, privacy,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2773600, IEEE Internet of Things Journal

lightweight cryptosystems and security protocols, software vulnerability, backdoor analysis, malware in IoT. Zhao et al. [69] surveyed security problems in the IoT in terms of node capture, fake node and malicious data, DoS attack, timing attack, routing threats, replay attack, side channel attack, mass node authentication problem, data access permissions, identity authentication, data protection and recovery, software vulnerabilities.

H. The Distribution of Publications in IoT

There are many studies and publications that can be categorized under the IoT subject. Table III summarizes the number of journal articles related to the IoT by year and subject. The number of journal articles in the table are collected by subject related search queries on Thomson Reuters Web of Science[11]. We further searched "learning" related journals on the same site with more specific search queries, such as popular learning techniques, related to the IoT, and these are listed in Table IV. Both tables show that the number of journal articles has been increasing in recent years in this area. It is also clear that the number of journal articles related to machine learning, deep learning, neural networks with IoT topics are also increasing recently.

III. Related Areas

The IoT is the end result of many developments in the past two decades. Here, we review related areas from a historical perspective. These include, but are not limited to, ubiquitous computing, pervasive computing, AmI, WSNs, WoT, SIoT, and Information Centric Networking. Figure 2 shows these areas and their development in time.

A. Ubiquitous Computing (UbiComp)

Ubiquitous computing is about computing anywhere, on any device, and in any format. UbiComp covers a wide range of study areas, including artificial intelligence, context aware computing, distributed computing, human-computer interaction, mobile computing, and sensor networks. Mark Weiser, who coined the phrase UbiComp, proposed to classify ubiquitous systems as "tabs" (wearable centimeter sized devices, such as smartphones), "pads" (hand-held decimeter-sized devices, such as laptops), and "boards" (meter sized interactive display devices, such as surface computers) [70][71]. However, this proposal does not include micro-size computing platforms. Therefore, it can be expanded to new forms: "dust" (miniaturized devices without display, e.g., micro electro-mechanical systems (MEMS)), "skin" (e.g., organic computer devices), and "clay"(ensembles of MEMS) [72].UbiComp is a transformation of real world objects to the virtual world nodes/objects. For instance, a smart meeting room that senses the existence of people in the room, records their actions and voices, obtains the writing from the whiteboard, and fuses the data of all sensors to extract meaningful knowledge [73].

[11]https://webofknowledge.com/

B. Pervasive Computing

The terms ubiquitous computing and pervasive computing are used interchangeably. However, they are conceptually different. UbiComp uses the advantages of mobile computing and pervasive computing. Mobile computing is expanding as technology improves and the number of computing devices increases. This enables mobile computing anywhere and anytime. However, it does not necessarily change or adapt the computing models based on context. On the contrary, pervasive computing treats context as a first-class citizen and adapts computing models based on context. Pervasive computing offers computing services that are invisible to the user. The objective of ubiquitous computing is to provide a pervasive computing environment when users change locations and context [74] [75].

Pervasive computing acquires context knowledge from the environment and provides dynamic, proactive, and context aware services to the user. Smart devices also provide location data, and therefore, context data. Not every smart device is suitable for applications requiring mobility due to the shape, weight, and battery power constraints. Smart devices can be adapted to the pervasive environment using the internet, web, and semantic web architectures [75].

C. Ambient Intelligence (AmI)

Ambient intelligence (AmI) is a concept in which the digital environment (sensor and device network) senses, computes interactions, and assists people in their daily lives. Ubiquitous and pervasive computing, sensors, networks, human-computer interfaces, and artificial intelligence are related to AmI; however, none of them individually cover AmI. AmI has invisible, intelligent, and flexible services that aim to benefit users and meet their expectations [76]. The European Commissions Information Society Technologies Advisory Group (ISTAG) supports the development of AmI with funding in the FP6 program [77][78]. ISTAG defines the components in AmI, including sensors, embedded devices, smart materials, MEMS, communication between devices, and adaptive software systems [79].

Smart homes, offices, buildings, and cities are among some of the examples that utilize AmI technologies. Embedded devices, sensors, actuators, and computers can communicate with each other using the network (sensor network, internet, web, etc.). Moreover, there are smart city IoT implementations used for daily life activities. SmartSantander[12] is an example of smart city IoT implementation. Authors of [80] also proposed smart city IoT implementation in their studies. Ambient assisted living (AAL) is another sample application area for AmI systems. AAL systems support elderly people in their daily lives and activities. Services and products (health, security, safety, mobility, social contact, etc.) that increase the quality of life consist of the primary applications of AAL [81]. There are also implementations of IoT platforms for AAL. Authors of [82] proposed a cloud-based IoT platform for AAL.

[12]http://smartsantander.eu/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2773600, IEEE Internet of Things Journal

TABLE III. NUMBER OF JOURNAL ARTICLES BY YEAR RELATED IOT AND SUBJECTS IN IOT

Subjects / Years	**2002**	**2004**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**Total**
”Internet of Things”	1	2	2	2	2	11	28	68	96	220	361	631	945	2369
”Artificial Intelligence” and “Internet of Things”	-	-	-	-	-	-	-	-	1	1	1	5	7	15
”Architectures” and “Internet of Things”	-	-	-	-	-	3	7	21	32	49	88	134	170	504
”Big Data” and “Internet of Things”	-	-	-	-	-	-	-	-	2	4	17	50	87	160
”Computation” and “Internet of Things”	-	-	-	-	-	-	-	2	12	17	36	80	62	209
”Context Awareness” and “Internet of Things”	-	-	-	-	-	-	-	-	4	14	25	31	19	93
”Machine Learning” and “Internet of Things”	-	-	-	-	-	-	-	-	-	-	1	6	14	21
”Security” and “Internet of Things”	-	-	-	-	1	4	7	13	16	41	69	121	165	436
”Semantic” and “Internet of Things”	-	-	-	-	-	-	1	3	7	13	29	21	32	106

TABLE IV. NUMBER OF JOURNAL ARTICLES BY YEAR RELATED LEARNING SUBJECTS IN IOT

Subjects / Years	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**Total**
”Artificial Intelligence” and “Internet of Things”	-	1	1	1	5	7	15
”Bayesian Network” and “Internet of Things”	-	1	1	2	-	-	4
”Context Awareness” and “Internet of Things”	-	4	14	25	31	19	93
”Deep Learning” and “Internet of Things”	-	-	-	-	1	7	8
”Expert System” and “Internet of Things”	-	1		1	-	1	3
”Fuzzy Logic” and “Internet of Things”	-	-	1	-	10	1	12
”Machine Learning” and “Internet of Things”	-	-	-	1	6	14	21
”Neural Network” and “Internet of Things”	1	2	2	2	4	10	21
”Support Vector Machine” and “Internet of Things”	-	-	1	2	1	-	4
”Decision Tree” and “Internet of Things”	-	-	-	-	-	2	2

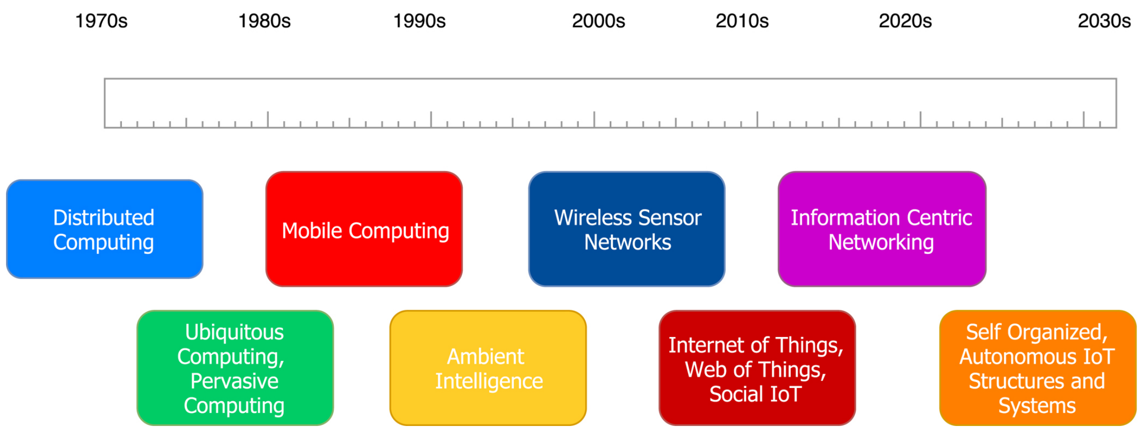

Fig. 2. IoT-related areas and their development in time.

D. Wireless Sensor Networks (WSN)

Low-power, low-cost, multifunctional sensor nodes can be combined and communicate with each other using wireless protocols. This is known as WSN. This network can be composed of hundreds or thousands of autonomous sensor nodes. Sensing, data processing, and communicating with each other are the tasks of these sensors in the network. Low-cost, energy-efficient, wireless, multi-hop, distributed sensing, and distributed computing are the characteristics and requirements of the WSN [83]. It is widely used in health, military, surveillance, computing, intelligence, control, reconnaissance, communications, and targeting systems. Wireless sensor standards are developed according to their power consumptions. IEEE 802.15.4, IEEE 802.15.3, ZigBee, WirelessHART, IETF, and 6LoWPAN are some of the standards [34] [84].

To overcome the complexity of network and data in WSN, semantic approaches were taken. Hence, semantic sensor networks were born.

1) Semantic Sensor Networks (SSN): Managing networks and data operations such as searching and querying are difficult tasks in sensor networks with complex structures and heterogeneity. Semantics help in declarative description of sensors, nodes, domains, and networks. SSNs provide these capabilities (managing, searching, and querying) with semantic definitions and reasoning over them. OWL (W3C Web Ontology Language) and RDF (Resource Description Framework) semantic languages are used in the definition of the SSN. Classifying the sensors with respect to their functionalities, output, and measurement, as well as inferring domain knowledge and classifying data according to spatially, temporally, accurately, producing events according to conditions, are potential capabilities of the SSN [85].

The W3C Semantic Sensor Network Incubator group (SSN-XG) developed SSN ontology that can describe sensors, relationships between sensors, stimulus, observations, accuracy, and the capabilities of sensors. SSN ontology can be seen in different perspectives: sensor, observation, system, feature, and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2773600, IEEE Internet of Things Journal

property. Sensor perspective focuses on "what senses, how it senses, and what is sensed". An observation perspective interests observation data. A system perspective focuses on "systems of sensors and deployments". A feature and property perspective is interested in "what senses a particular property or what observations have been made about a property" [86].

2) Semantic Sensor Web (SSW): When sensors are connected to the internet and web, they become part of what is called a semantic sensor web (SSW). Sensors in a SSW can then discover new sensors, and share semantic sensor data (time stamp and spatial coordinates) with each other. The Open Geospatial Consortium (OGC) and semantic web activity of the W3C enhance the SSN and standardize the SSW. The sensor web is a specialized area that has a web-centric information infrastructure (collecting, modeling, reasoning, storing). The SSW uses ontologies, rules, semantic languages OWL and RDF for interoperability, reasoning, and analysis of sensor data from different devices and platforms [8] [87].

E. Web of Things (WoT)

The WoT is a specialized area of the IoT that uses standard web technologies in all areas. Web technologies include representational state transfer (REST), hyper-text transfer protocol (HTTP), transmission control protocol (TCP), user datagram protocol (UDP), Internet protocol (IP), JavaScript object notation (JSON), JavaScript object notation for linked data (JSON-LD), microdata, and Web sockets for intercommunication, data processing, and visualization [7]. Sensors and devices in WoT can be connected to the web directly and their data can be processed in the cloud. Therefore, it is larger than SSW.

F. Semantic Web of Things (SWoT)

The SWoT is an intersection area that combines semantic technologies and IoT. It is an evolution of WoT. With semantic technologies, WoT sensors and devices can be defined semantically with semantic languages, such as OWL and RDF, that support reasoning, storing (triple store), querying (with SPARQL RDF Query Language), searching, and monitoring [9].

G. Social Internet of Things (SIoT)

The SIoT is a term coined to signify that sensors and devices are connected with each other and with humans via specialized social networks [36]. Holmquist et al. [88] initially proposed the idea. In their proposal, smart devices used WSN as a social network. With the development of the IoT, devices can use internet instead of WSN. Atzori et al. [46] proposed the SIoT for the interacting network of sensors and devices as human social networks. With the SIoT, devices can be discoverable and reachable. There are different SIoT platforms and implementations. The SWoT, Evrythng, Paraimpu, Xively, and Toyota Friend Network (automobile WoT social network) are some of them [35].

H. Information Centric Networking

Information Centric Networking (ICN) is a type of Internet architecture that differs from the IP address-centric model. It can be considered as an Internet structure to be implemented in the future, for which the development is underway. In the ICN approach, data is independent of server location, distribution channel, and application. Content is reachable with unique ICN names. ICN architecture has the following features: "In-network caching, content-based naming and security, name-based content discovery and delivery, and a connectionless receiver-driven communication model" [53]. There are different ICN approaches, namely data-oriented, content-centric, publish-subscribe network architecture, and network of information. These design models are compared in [89].

In addition, ICN can be used in the IoT world by considering IoT data as content. Without using the IP request-reply mechanism, IoT sensor data can be reachable with addressable content by using ICN, which is helpful for mapping the digital world to the physical world [53]. Moreover, using ICN is beneficial in fulfilling the general requirements of the IoT. Energy efficiency, heterogeneity, mobility, quality of service, scalability, and security IoT requirements are handled with ICN features. These are known as anycasting, content-based security, connection-less mode, data, in-network caching, and multicasting. Amadeo et al. [53] discussed and summarized these in their paper.

I. Software Defined Networking

Software Defined Networking (SDN) and Network Virtualization (NV) are two technologies that can assist in the solution of the fundamental IoT problems such as scalability, interoperability of heteregeneous devices, discoverability, security, management and application specific requirements. With the usage of SDN and NV, IoT networks can be more dynamic, elastic and scalable [90]. In literature, there are different papers that mentioned IoT-SDN-NV relations. Bizanis et al. [91] surveyed SDN and NV solutions for the IoT devices. They mentioned SDN and virtualization solutions for mobile and cellular networks, wireless sensor networks and IoT architectures. SDN based, Virtualization based IoT architectures and SDN based important frameworks such as UBiFlow, SDIoT, MINA are mentioned in this survey paper. There are also novel studies and researches about SDN-IoT architectures. Qin et al. [92] implemented the extended MINA-SDN prototype for IoT scenarios such as integration of the electric vehicles, electric charging sites, smart grid infrastructures. Bedhief et al. [93] proposed a SDN-Docker based IoT architecture that consists of Dockers implemented IoT devices.

IV. Context-Awareness

This section discusses the definition of context awareness, features, and levels of context awareness, context awareness design principles, context life cycle (context acquisition, modeling, reasoning, and distribution) and context aware solutions and technology.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2773600, IEEE Internet of Things Journal

A. What are Context and Context Awareness?

Various definitions of context exist. The word context is defined as "ambience, attitude, circumstance, dependence, environment, location, occasion, perspective, phase, place, position, posture, situation, status, standing, surroundings and terms" in dictionaries (Thesaurus[13]). In literature, Abowd et al. [94] [95], Sanchez et al. [96] defined context as: "any information that can be used to characterise the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves [94]." Abowd et al. [95] also identified the minimum requirements - five W's (what, who, where, why and when) in order to be able to analyze and understand context. In addition to these definitions, Sanchez et al. explained the difference between raw data (unprocessed, directly from sources) and context information (processed, generated from raw data).

B. Context Aware Features and Context Types

Perera et al. [21] defined the features of context awareness related with IoT, according to Abowd et al.[94], Schilit et al.[10], Pascoe [97]. Presentation, execution, and tagging (annotation) were the features of context aware systems. Within that concept, a context aware system decided what information and which service should be presented to the user. When the user went for shopping, the system perceived that user during the shopping process and the shopping list was updated according to the smart fridge decision. This feature was called presentation. In an IoT system, devices should execute automatically. When the user has a health problem, such as a heart attack, the context aware system has to consider the possible problems with the corresponding wearable IoT devices, and if necessary, call the nearest hospital for emergency, sending the user's location to the smart car for fast transportation. With this feature, the devices in the system should execute automatically and synchronously. Tagging (annotation) is another feature of a context aware system. Single sensor data might not be necessary for interpretation and analysis. As a result, the sensor data produced by multiple sensors can be used for interpretation and analysis. Multiple sensor data fusion is an important step for capturing meaningful information. Tagging, also known as context annotation, provides particular information about which data is associated with which sensor and what value is stored for that sensor. Sensor location and data collection time are among the other issues that should be tagged for context awareness.

C. Context Life Cycle in IoT

The context life cycle expresses how sensor data is collected, modeled, and processed, and how knowledge is extracted from the collected data. As a result, it is beneficial to develop frameworks, structures, and solutions for the IoT. Perara et al. [21] mentioned that there were different context life cycles. Common characteristics of these life cycles are categorized into four main parts: context acquisition, context modeling, context reasoning, and context distribution [21] [26]. In the context acquisition phase, the data is acquired from various physical and virtual sensors. In the context modeling part, the data is required to be modeled according to the meaningful data. In the context reasoning phase, raw data is required to be processed first, and then, the knowledge is extracted. For the last category, context distribution, the obtained knowledge is distributed via servers, query languages, and frameworks. Figure 3 illustrates Context life cycle in IoT.

1) Context Acquisition: Context acquisition is evaluated based on five factors: the acquisition process, frequency, responsibility, sensor types, and source. These five factors are explained in [21].

2) Context Modelling: Context modeling is defined as the context representation that provides assistance in the understanding of properties, relationship, and details of context. It varies depending on the domain and features of the context. Context modeling varies depending on the requirements that are defined as distributed composition, efficient context provisioning, dependencies and relationships, heterogeneity and mobility, imperfection, incompleteness and ambiguity, level of formality, partial validation, reasoning, richness and quality of information, timeliness, and usability of modeling formalisms [98] [23]. The most popular classification of context modeling is defined as key-value, markup scheme, graphical, object oriented, logic based, ontology based, spatial, uncertainty, and hybrid context modeling [98],[21],[99] [23] [26]. There are also different approaches for context modeling, such as multidisciplinary, domain focused, user-centric, and chemistry inspired [26].

- Key-Value Modeling: In this model, values are stored as key-value pairs that provide simplicity, flexibility and user-friendliness. However, this particular modeling method is not convenient for complex and hierarchical structures and relationships. It is hard to retrieve information and there is no standard tool for processing. Meanwhile, it is suitable for temporal storage, application preferences, and configuration.
- Markup Scheme Modeling: In this model, tagging is used to store data. XML is probably the most frequently used markup language for markup scheme modeling that provides temporarily data storage. However, it does not support reasoning, and retrieval of information is difficult. Composite capabilities/preference profiles (CC/PP) [14] is another commonly used markup scheme model.
- Graphical Modeling: In this modeling technique, the context is modeled with relationships. Unified modeling language (UML)[15] and and object role modeling (ORM)[16] represent some of the examples of this modeling technique. Relational SQL, NoSql databases, and XML can be used for graphical modeling. This modeling is better than key-value and markup scheme modeling in terms of modeling context.

[13]http://www.thesaurus.com/

[14]https://www.w3.org/TR/2007/WD-CCPP-struct-vocab2-20070430/

[15]uml.org

[16]ormfoundation.org

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2773600, IEEE Internet of Things Journal

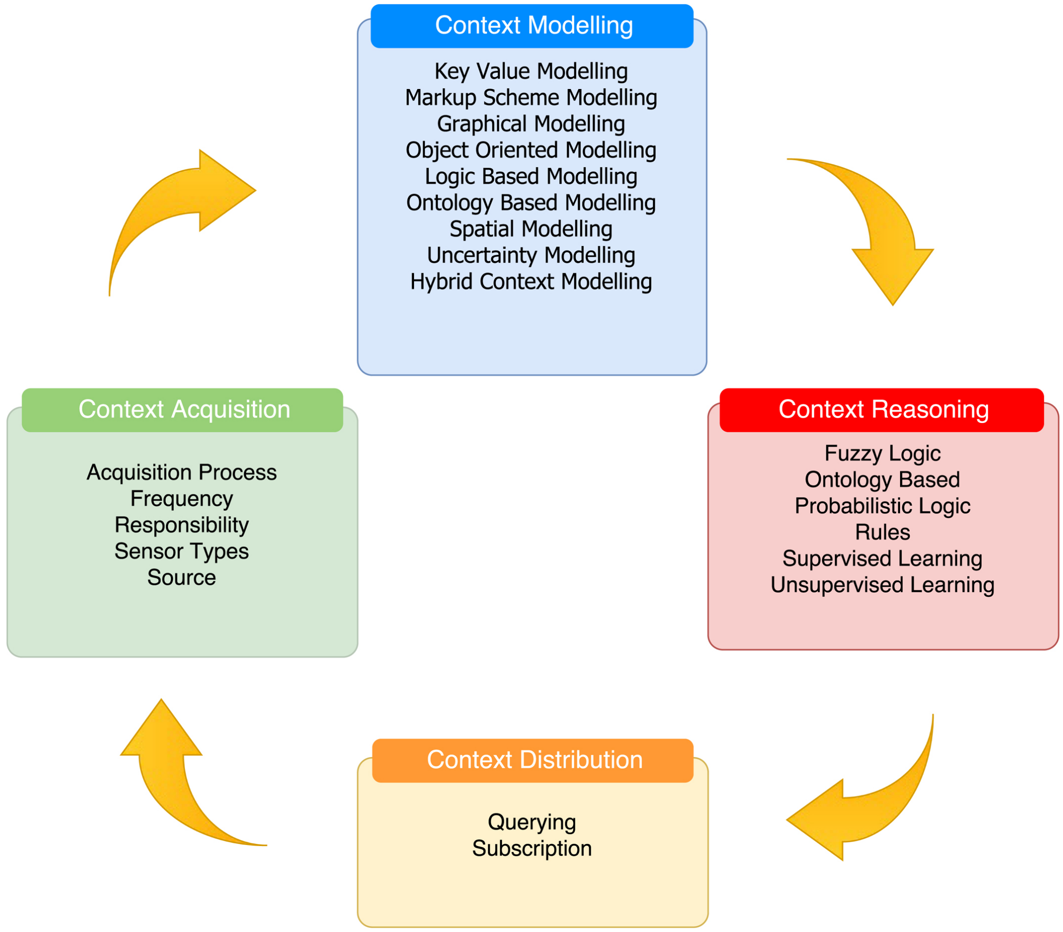

Fig. 3. Context Life Cycle in IoT

- Object Oriented Modeling: In object oriented modeling, classes and relationships are used for modeling context and data. Object oriented, high level languages are suitable for this modeling. However, validation is difficult due to specifications and standardizations. In addition, in this modeling, reasoning is not convenient for inference.
- Logic-Based Modeling: In this model, context is represented with rules, logical expressions, and variables. Logical reasoning can be applied with an implementation of logic-based modeling. Additionally, high-level context can be extracted with low-level context, which can be provided via existing processing tools. However, validation and standardization are difficult to implement in this model.
- Ontology Based Modeling: In ontology based modeling, context is modeled with ontology, and represented with semantic ontology languages, such as OWL, RDF, and RDF Schema (RDFS). These languages are used in semantic technologies (semantic web, databases, Sparql, etc.) With ontology based modeling, both reasoning and knowledge extraction can be implemented. In addition, rich context expressions and strong validation can be provided. However, information retrieval may be complicated due to the complex representation of ontology languages. In Section VI, ontology definition, ontology languages, and review of the papers related to semantic technologies will be covered.
- Spatial Modeling: In this model, physical space, location of sensors, and real world entities are modeled and expressed as context information. Geometric (latitude, longitude, elevation, etc.) and symbolic (room number, ID of access point, etc.) coordinates are used by positioning systems for spatial modeling. Spatial context modeling is achieved with tiers of spatial ontologies: tier0 (ontology of the physical reality), tier1 (observations of reality), tier2 (observations are defined by uniform properties), tier3 (social reality and relations of all objects) and tier4 (rules are modeled) [100].
- Uncertainty Modeling: The physical world contains uncertainties and ambiguities. Meanwhile, real world modeling also causes imperfection and ambiguity issues. Therefore, in order to solve that particular problem, quality of context is proposed by researchers. Quality of context can be measured in terms of attributes, such

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2773600, IEEE Internet of Things Journal

as accuracy, coverage, confidence, frequency, freshness, repeatability, resolution, and timeliness [101] [102].

- Hybrid Context Modeling: In this model, combined and hybrid modeling techniques are used. Hybrid fact-based/ontological modeling and markup-based/ontological modeling are some examples of hybrid context modeling [103] [104]. In hybrid modeling, the advantages of modeling techniques are used. For instance, ontological modeling is used for rule-based reasoning.

3) Context Reasoning: Context reasoning can be described as the extraction of new knowledge from the available context and extraction of context sets from high level context for better understanding [105] [106]. Uncertainty and imperfection of raw data are the requirements for context reasoning. There are three main steps of context reasoning: context preprocessing, sensor fusion, and context inference [107] [21]. In the context pre-processing phase, data are cleaned and formed by defining the relevant context attributes (dimensionality reduction, feature subset induction), filling missing data, validating context, removing outliers, and using data mining techniques for preparation of processing. In sensor fusion step, multiple sensor data are combined to produce more dependable, accurate, reliable, and complete data that cannot be provided by single sensor data. In the inference phase, recognizing new context, which is relevant, and mapping lower level context to higher level context (logical and probabilistic reasoning) are important steps in producing high level context for inference. Reasoning approaches for different context awareness problems are listed in Table V.

There are also different context reasoning and inference models, in terms of artificial neural networks, decision tree, fuzzy reasoning, hidden Markov models, k-nearest neighbor, naive Bayes, ontology-based, rule-based, support vector machines, etc. In Section V, these learning schemes, decision techniques and related papers will be compared and surveyed in the IoT perspective. However, in the following section, main methods for context reasoning will be mentioned. Context reasoning can be classified as different categories [21]: fuzzy logic, ontology-based, probabilistic logic, rules, supervised learning, and unsupervised learning. Some of the important papers and solutions are listed and compared in Table VI.

- Fuzzy logic: Fuzzy logic is different from traditional logic. In traditional logic, everything is represented with 0 or 1. However, in fuzzy logic, partial truth is also acceptable. In this way, the representation of the real world with fuzzy logic is more acceptable than using traditional logic (speed: slightly fast, very slow, etc.). Fuzzy logic reasoning technique is not used standalone. Instead, it is most frequently used with other reasoning techniques with regard to ontological, probabilistic and rule-based reasoning.
- Ontology-based: Ontology-based logic depends on description logic, and reasoning can be achieved with ontology modeled data. Semantic web languages, such as RDF, RDFS, and OWL are used to implement ontology-based reasoning. It can be combined with ontology modeling, which is the advantage of this reasoning. However, it is not capable of supplying missing values and finding ambiguity. Thus, it is mostly used with rule-based reasoning. Event detection and hybrid reasoning are some examples of the application domain of ontology-based reasoning.
- Probabilistic logic: In this technique, decisions are based on the calculation of event probabilities and facts. Different sensor data are combined with probabilistic logic. Dempster-Shafer and Hidden Markov Models are used as probabilistic reasoning for predicting the next event, recognizing activities, and forecasting uncertainty. Dempster-Shafer uses sensor data fusion to calculate the probability of events. Hidden Markov Models provide vision for the next state by using the current state. They are most frequently used in context awareness.
- Rules: In this technique, reasoning can be acquired with an If-Else structure. Rules are used with ontological reasoning. User preferences, event detection, and human thought can be modeled with rules for use in IoT applications.
- Supervised learning: Generally, in this technique, sensor data are collected and labeled for training. Then, functions and algorithms are generated according to the expected data, and they are applied to all available data. An ANN is a supervised learning technique used for finding patterns and modeling complex models between input and output. The Bayesian network is another technique that is used for probabilistic reasoning. Directed acyclic graphs are used in Bayesian networks to express events and relationships. The decision tree is also used as a supervised technique to build a tree for classification of data. Support vector machines are used for recognizing patterns. In Section V, related papers and solutions will be reviewed and compared.
- Unsupervised learning: In this technique, clustering is used to extract meaningful results from unlabeled data. The k-nearest neighbor clustering technique is used for context aware reasoning. Low-level, simple actions and operations (positioning and location) can be resolved with a clustering technique. Another unsupervised learning technique is the Kohonen self-organizing map (KSOM) (unsupervised neural network technique) that is used for classifying incoming real sensor data and for context aware applications, such as noise and outlier detection [21].

4) Context Distribution: Context distribution delivers context to users. Context acquisition methods can also be context distribution methods, in terms of the user's changing usage perspective [21]. There are other context distribution techniques in regards to querying and subscription. In the querying method, users create queries to produce results. In the subscription method, users subscribe to the context system to periodically obtain specific sensor data, or when a specific event occurs. This method is used for real-time processing systems.

V. Machine Learning

In this section, we focus on machine learning algorithms, which are used in areas such as ubiquitous computing, AmI,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2773600, IEEE Internet of Things Journal

TABLE V. REASONING APPROACHES FOR CONTEXT AWARENESS: ADAPTED FROM [108]

Context	Reasoning Approaches
Activity ((Driving, Eating, Running, Sitting, Sleeping, Speaking, Walking)	Hidden Markov Models, Decision Tree, Bayesian Networks
Availability(Busy, Free)	Rules, Decision Tree
Environmental (temperature, humidity, raining, etc.)	Direct from sensors
High Level Identity(family member, neighbor, friend, etc.)	Social Ontology, Rules
High Level Location(work, office, school, home, etc.)	Bayesian network, Decision Tree, Rules, Logic
Low Level Identity	Specific Database Lookup
Low Level Location (Coordinates, GPS)	Direct from device
Mobility	Hidden Markov Models, Dynamic Bayesian Networks
Proximity(close, near, far)	Fuzzy Logic
Physiological Biometric (Blood Glucose, Oxygen Saturation, Pulse, etc.)	Direct from Bio Sensors
Social Context (What's their availability?, Who is nearby?)	Logic, Rules
Temporal (Time, day, year)	Direct from time sources

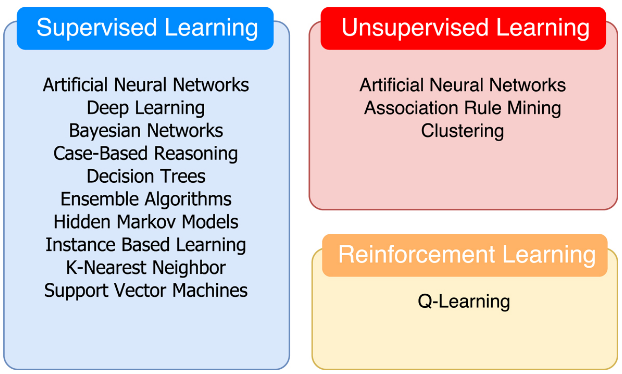

Fig. 4. IoT Related Machine Learning Algorithms

pervasive computing, mobile computing, context aware systems, and sensor networks that are all related to IoT. We review supervised learning, unsupervised learning, and reinforcement learning techniques in the following subsections. Since machine learning has a very wide coverage span, we review only IoT related and context aware machine learning-related algorithms and techniques here. Figure 4 illustrates IoT related (and also general) machine learning algorithms and their classifications.

A. *Supervised Learning*

In supervised learning training data, which consists of a set of samples of labeled data, is used to learn and train a model. Then this model is later used to predict new sample data. There are different supervised learning techniques. Main techniques that are relevant to context reasoning in IoT can be listed as follows: artificial neural networks, Bayesian networks, case-based reasoning, decision trees, ensembles of classifiers, hidden Markov models, instance-based learning, k-nearest neighbor, and support vector machines. Some of the important papers that are focused on supervised learning algorithms are listed and compared in Table VII.

1) Artificial Neural Networks (ANN): ANN mimics a biological neural network, which provides an autonomous learning structure. There are different types of ANNs in the supervised learning paradigm: backpropagation, ensemble, multilayer perceptron, Hopfield networks, and Boltzmann machines. There are also other ANN types in unsupervised and reinforcement learning paradigm. Backpropagation neural networks are used in human activity recognition (walking, running, sitting, etc.) with wearable sensors [139] [140]. Guan et al. [105] used 40 sensors for both legs to sense activity recognition and used backpropagation neural networks for context reasoning. Saeedi et al. [141] used smartphone sensors (accelerometer, gyroscope, GPS, magnetometer, and temperature) for personal navigation that used backpropagation neural networks for the learning part of the system. In addition, Choi et al. [142] used neural networks for the learning section of a smart home application system. Moreover, Mishra et al. [143] used multilayer perceptron and a fuzzy-neuro genetic algorithm to process IoT big data in their framework, which is a cognitive-oriented IoT big data framework.

2) Deep Learning: Deep learning is a type of ANN that consists of multiple processing layers and enables high level abstraction to model data [144]. There are different types of deep learning models, namely, backpropagation (with multiple hidden layers), Convolutional Deep Neural Networks (CNN), Recurrent Neural Networks (RNN), deep belief networks, restricted Boltzmann machines (RBMs), and long short term memory (LSTM) networks.

In the IoT world, deep learning techniques have been used in recent years. In literature, the use and implementation of deep learning techniques for learning from data collected in the IoT and cloud are still under development. Lane et al. [145] researched the use of deep learning implementations to process data from IoT devices, such as wearables and smart phones. They compared IoT hardware platforms (Snapdragon 800, Tegra K1, Edison) in terms of energy consumption, execution time, and other performance metrics, when implementing DNN and CNN models to process audio and image sensor data. De Coninck et al. [146] proposed The Big-Little Approach, implementing DNN in IoT. In their approach, if the little DNN in a smart device cannot classify input data, this input data is sent to a remote cloud DNN for classification. Ma et al. [147] used an RNN and deep restricted Boltzmann machine to fit a model and predict traffic congestion in China using GPS data from taxis. Zhang et al. [148] used a deep autoencoder to process fused IoT data (sensor, social, and background data) for pollution detection and traffic patterns. Alsheikh et al. [149] proposed analysis of mobile (human activity recognition) data using a deep learning neural network. They used Apache Spark to process mobile big data in parallel. In addition, deep learning algorithms are used for feature extraction from sensor data. Plötz et al.[150] used deep belief networks, RBMs

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2773600, IEEE Internet of Things Journal

TABLE VI. Context Reasoning Techniques and Related Papers

References	Fuzzy Logic	Ontology Based	Probabilistic	Rules	S.Learning	U.Learning	Logic
Ranganathan et al. [109]	✓						
Mantyjarvi et al. [110]	✓						
Padovitz et al.[111]	✓		✓				
Bikakis et al.[106]		✓					
Teymourian et al.[112]		✓		✓			
Song et al.[113]		✓					
Zafeiropoulos et al. [114]		✓					
Zafeiropoulos et al. [115]		✓					
Liu et al. [116]			✓				
Zhang et al. [117]			✓				
Keßler et al. [118]				✓			
Choi et al. [119]				✓			
Barbero et al. [120]				✓			
Konstantinou et al. [121]				✓			
Terada et al. [122]				✓			
Gyrard [9]		✓		✓			
Lane et al.[33]					✓		
Riboni et al. [123]		✓			✓		
Korel et al. [24]					✓	✓	
Huang et al. [124]					✓		
Doukas et al.[125]					✓		
Park et al.[126]					✓		
Brdiczka et al.[127]			✓		✓		
Tsung-NanLin et al. [128]						✓	
VanLaerhoven et al. [129]						✓	
Shtykh et al. [130]						✓	
Chihani [131]		✓		✓			
Perera et al. [2]		✓		✓			
Ranganathan et al. [132]		✓					✓
Preuveneers et al. [133]		✓					
Serral et al. [134]		✓					
Beamon et al. [108]		✓		✓			✓
Hong et al. [135]		✓			✓		
Hagras et al. [136]	✓			✓			
Riboni et al. [137]		✓					✓
Guan et al. [105]				✓	✓		
Jie et al. [138]				✓	✓		

and compare the performance of deep learning algorithms with statistical and fast fourier transform (FFT) based feature extraction approaches.

3) Bayesian Networks: Bayesian networks are the types of networks based on directed acyclic graphs. In this method, sensor data are classified with random variables and statistics. In addition, a Naive Bayes (NB) classifier is a specific type of classifier based on the assumption that the representative features for classification might be considered independent. Bayesian networks and NB classifiers are used in human activity recognition [151] and health monitoring areas, such as monitoring and classifying respiration rate, step count [160] and heart rate data [152] [153]. In addition to these, Korpipää et al. [154] used Bayesian networks and NB classifiers to extract features from audio data and to classify data. In their studies, they recognized activities according to context. Ramakrishnan et al. [155] used Bayesian networks to implement a heterarchical autonomic recursive distributed Bayesian network (HARD-BN) framework that provides learning and reasoning using IoT sensor data. Further, Krause et al. [156] used Bayesian networks and NB classifiers to implement wearable sensor platform probabilistic reasoning parts that support inference from collecting data using sensors, such as armbands, headsets, smart phones, and GPS receivers. In addition to these, Bayesian networks and NB classifiers are also commonly used solutions and learning models for home automation [157] and navigation systems [141].

4) Case-based Reasoning: In this supervised learning method, new problems are solved using past solutions of similar problems. It is similar to making analogies. There are four main steps for case-based reasoning: retrieve, reuse, revise, and retain [158]. A past problem, which is similar to the new problem, is retrieved from memory. The solution of that problem is reused for the new problem. Problem attributes are revised according to the new case, and the solution is retained in memory. Paper [159] is one example of case-based reasoning.

5) Decision Trees: In this method, values and attributes are implemented with tree hierarchical models and mapped with nodes and edges. Classification rules are used in traversing from root to leaf. The C4.5 algorithm (extension of ID3) is a widely used decision tree algorithm [30] [151]. Human

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2773600, IEEE Internet of Things Journal

TABLE VII. SUPERVISED LEARNING ALGORITHMS - ARTIFICIAL NEURAL NETWORKS (ANN), DEEP NEURAL NETWORKS (DNN), BAYESIAN NETWORKS (BN), CASE-BASED REASONING (CBR), DECISION TREES (DT), ENSEMBLES OF CLASSIFIERS (EC), HIDDEN MARKOV MODELS (HMM), INSTANCE-BASED LEARNING (IBL), K-NEAREST NEIGHBOR (KNN) AND SUPPORT VECTOR MACHINES (SVM)

References	ANN	DNN	BN	CBR	DT	EC	HMM	IBL	KNN	SVM
Khan et al. [139]	✓									
Altun et al. [140]	✓								✓	
Guan et al. [105]	✓								✓	
Saeedi et al. [141]	✓		✓							
JonghwaChoi et al. [142]	✓									
Bao et al. [151]			✓		✓					
Tapia et al. [152]			✓							
Lara et al. [153]			✓							
Korpipää et al. [154]			✓							
Ramakrishnan et al. [155]			✓							
Krause et al. [156]			✓							
Bhide et al. [157]			✓							
Aamodt et al. [158]				✓						
TinghuaiMa et al. [159]				✓						
Lara et al. [30]					✓					
Jatoba et al. [160]					✓				✓	
Ermes et al. [161]					✓					
Huebscher et al. [162]					✓					
Byun et al. [163]					✓					
Brdiczka et al. [164]					✓					
Alpaydin et al. [165]						✓				
Anagnostopoulos et al. [166]						✓				
Ravi et al. [167]			✓			✓			✓	✓
Mannini et al. [168]	✓		✓				✓		✓	✓
Oliver et al. [169]							✓			
Chen et al. [170]			✓				✓		✓	✓
Zhen-YuHe et al. [171]										✓
Kranz et al. [172]										✓
Khan et al. [173]										✓
Schmitt et al. [174]	✓		✓		✓					✓
Hong et al. [135]					✓					
Sasidharan et al. [175]	✓									
Lane et al. [145]	✓									
Mishra et al. [143]	✓									
Lane et al. [145]		✓								
De Coninck et al. [146]		✓								
Ma et al. [147]		✓								
Zhang et al. [148]		✓								
Alsheikh et al. [149]		✓								

TABLE VIII. UNSUPERVISED LEARNING ALGORITHMS - KOHONENS SELF-ORGANIZING MAP (SOM) , THE RECURRENT SELF-ORGANIZING MAP (RSOM), NEURAL GAS (NG), ASSOCIATION RULE LEARNING (ARL)-APPRIORI, K-MEANS, DBSCAN (DENSITY BASED SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE)

References	SOM	RSOM	NG	ARL	KM	DBSCAN
Van Laerhoven et al. [176]	✓	✓			✓	
Van Laerhoven et al. [177]	✓	✓			✓	
Mayrhofer et al. [178]	✓	✓	✓			✓
Guo et al. [179]				✓		
Ramakrishnan et al. [155]				✓		

activity recognition papers Bao et al. [151], Jatoba et al. [160] (used CART, ID3 algorithms), Ermes et al. [161] are among the examples using decision trees algorithms. In these papers, human activities (walking, running, sitting, etc.) are determined with wearable sensor networks (acceleration sensors, etc.) and embedded devices. In addition, Huebscher et al. [162] used a relevance-based decision tree learning approach to implement an adaptive middleware framework for context aware applications. Moreover, Byun et al. [163] proposed a solution that provides a dynamic adaptation using decision tree algorithms. Brdiczka et al. [164] proposed a context model for an intelligent environment that used decision trees algorithms for reasoning.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2773600, IEEE Internet of Things Journal

6) Ensemble Algorithms (Ensembles of Classifiers): In this method, multiple classifier algorithms and models are used together to achieve better classification, instead of using a single classifier and model. There are different methods of ensembles of classifiers: voting, bagging, and boosting. In the voting method, each classifier returns a result. The final classifier is generated with the greatest number of results, that is called voting. In the bagging method, subsets of the initial training set are formed by random sampling with replacement. The base classifier is generated (learned) using each sub-sample. In the boosting method, sub-samples of the training set are weighted according to difficulty of the classification. The boosting method is similar to the bagging method. However, the learning for the final classifier is achieved through the weighted votes of the classifiers [165] [166]. Anagnostopoulos et al. [166] used AdaBoost M1, bagging, voting methods, and naive Bayes classifiers in their proposed model, for predicting the location of mobile users. Ravi et al. [167] used decision trees, k-nearest neighbors, SVM, and naive Bayes algorithms with voting methods in their study, Activity Recognition from Accelerometer Data.

7) Hidden Markov Models (HMM): Hidden Markov Models are simple dynamic Bayesian networks based on the statistical Markov model (probabilistic mathematical model that defines future states depending on the current state). The system is modeled with unobserved hidden layers. HMM is used for the context reasoning and learning part of the IoT systems. Manini et al. [168] used HMM for human physical activity reasoning in their studies. On-body accelerometers collect sensor data and classify data with ANN, KNN, HMM, naive Bayes, and SVM. Oliver et al. [169] used HMM to determine office activities from multiple sensory channels. Chen et al. [170] used the Bayesian network, SVM, KNN, and HMM in a proposed context-aware search system for the IoT. In addition to these, Chen et al. [180] used Markov chains to implement a collaborative sensing intelligence framework in an industrial IoT.

8) Instance-Based Learning : In this technique, new instances are compared with other instances in training sets. Comparison can be achieved with a distance function to determine the distances of each instance pair causing expensive results, such as more memory requirements and computation time. Thus, the learning part should not be implemented on the mobile side/device. The cloud side is more convenient for storage and computing. The k-nearest neighbor is one example of instance-based learning.

9) K-Nearest Neighbor (KNN): KNN is a type of supervised learning algorithm, such that the values are classified with majority voting. In this method, new values are categorized with the greatest number of nearest neighbor values. The KNN algorithm is used in human activity recognition and context aware systems. Altun et al. [140] proposed a system that contains inertial and magnetic sensor units that sense human activity. In their study, they used the KNN, Bayesian decision making (BDM), least-squares method (LSM), dynamic time warping (DTW), support vector machines (SVM), and ANN. Guan et. al. [105] used KNN and backpropagation neural networks for context reasoning. They were able to achieve over 90% reasoning success. In addition, Jatoba et al. [160] used the KNN for their experiments.

10)Support Vector Machines (SVM): The SVM is a popular supervised classifier and pattern recognition algorithm that provides data analysis for classification and regression. Classification can be achieved with support vectors, which provide the optimum distance of target class boundaries. The SVM is used in different areas, such as image, video processing, pattern recognition, sensor data analysis, and recognition. Zhen-YuHe et al. [171] used a SVM in a human activity recognition model to determine the activities of humans. Kranz et al. [172] used a SVM in their middleware that collects data from sensors and classifies the data for reasoning. Khan et al. [173] proposed a method that classifies streaming data from IoT devices by using simple aggregation approximation, density-based clustering and a SVM.

B. Unsupervised Learning

The aim of the unsupervised learning technique is to cluster the unlabeled data. In this technique, there is no classified and labeled data. Results are returned faster than the supervised learning approach for hidden patterns and big data. Further, in this technique, a large stack of heterogeneous data is divided into easily understandable and manageable smaller homogeneous subsets. There are different unsupervised learning approaches: ANN, association rule learning, and clustering. Some important papers focused on unsupervised learning algorithms are listed and compared in Table VIII.

1) Artificial Neural Network: Self-organizing map (SOM) is a type of unsupervised learning algorithm that uses ANN. In an SOM algorithm, high dimensional spaces are mapped to neurons that store the data. It is a type of clustering, vector quantization algorithm. There are different types of SOM: Kohonens self-organizing map (KSOM), and recurrent self-organizing map (RSOM). Neural gas is a type of ANN that is based on SOM. Finding optimum values, which are based on the feature vector, can be achieved with neural gas.

Van Laerhoven et al. [176] [177] used KSOM, RSOM, k-means and Hartigans sequential leader clustering algorithms to analyze real-time sensor data from wearable sensors for comparison of unsupervised ANN algorithms. Further, Van Laerhoven et al. [176] used SOM algorithms for teaching context to algorithms. In addition, Mayrhofer et al. [178] used SOM, RSOM and k-means Hartigans sequential leader clustering, growing k-means clustering, neural gas, neural gas with competitive Hebbian learning (NG+CHL), growing neural gas (GNG), and incremental DBSCAN (IDBSCAN), to recognize and predict context by learning from user behavior by using sensors.

2) Association Rule Learning: The aim of this method is to find interesting relations between variables. There are different association rule learning algorithms, such as apriori, eclat and FP-growth. Guo et al. [179] used the apriori algorithm and map reduce model in IoT cloud computing datasets to mine frequent structures. In paper [155], an adapted apriori algorithm is used for finding correlations.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2773600, IEEE Internet of Things Journal

3) Clustering: There are different clustering algorithms in unsupervised learning techniques: K-means algorithm, fuzzy clustering, DBSCAN (Density based spatial clustering of applications with noise), OPTICS (Ordering points to identify the clustering structure) algorithm. K-means is the most frequently used clustering algorithm, which provides the minimum distance between similar data, and maximum difference between clusters. In the k-means clustering algorithm, N input datasets divide the cluster sets into k pieces. Mayrhofer et al. [178] used the k-means clustering algorithm for context reasoning. Van Laerhoven et al. [176] also used the k-means algorithm to cluster real-time wearable sensor data.

C. Reinforcement Learning

Reinforcement Learning (RL) is another type of learning approach used in various areas, such as control theory, simulation-based optimization, statics, and other fields related to automatic control. In the RL approach, automatic control (convergence to the desired state and ideal behavior) is provided with a feedback signal, which is called the reinforcement signal. In context awareness studies in WSNs, RL is used to improve the system performance. The features of the RL are examined under the following subtitles: state representation, event representation, action representation, rule representation, reward representation, agent interaction and control, and exploration versus exploitation [181].

Q-Learning is the most commonly used RL approach in context awareness of the WSN [182]. In Q-Learning, the RL system chooses a subset of actions, which are defined in rules, then, it determines exploration (random action) or exploitation (choosing action with best Q-value from Q-table). After execution of the action, the state, event, and reward are observed. According to the state of the environment, the Q-table and rules are updated [181].

VI. Big Data

In this section, we are focusing on data issues related to the IoT. Due to the rapid growth in IoT installations, the corresponding data problems associated with big data are becoming more common. Here, the definition and features of big data, big data generation and acquisition, big data storage, and finally big data analytics will be reviewed in the context of IoT. Figure 5 presents IoT big data-related research and development areas.

A. Definition and Features of Big Data

Big data refers to very large unstructured data as compared to other types of datasets. Hidden and new information can be found when analyzing this big data, and that is why it is becoming more interesting and receives significant attention from both industry and academics. In the IoT, sensors create significant data and this will rapidly increase. For example, the number of RFID tags are expected to reach 209 billion by 2021 [37]. The number of connected IoT devices are estimated to be in the range of 50 to 100 billion in very near future [2], and the total amount of data on Earth will reach 35 zettabytes (ZB) soon [11]. These numbers show that IoT data will be part of the big data.

Big data differs from the traditional data in terms of "volume" (great volume and collection), "variety" (various types of data), "velocity" (rapid production, generation and analysis), "value" (low density and huge value), "variability" (inconsistency of data prevents easy handling and processing), and "veracity" (quality of data vary greatly) [37] [183]. Traditional computers and systems were not sufficient to process, analyze, and manage big data when Internet companies, such as Google, Yahoo, and Facebook, started collecting huge amounts of data in the early 2000s. Thus, they developed new systems, new approaches, and solutions to process big data.

IoT big data has different characteristics when compared to common big data problems. These characteristics can be classified into three categories, namely, data generation, data quality, and data interoperability [32]. In data generation, velocity (generation of data at different rates), scalability, dynamics (things are mostly mobile in IoT), and heterogeneity (different types of sensors) are specific characteristics that make IoT data different from others. From data quality perspective, uncertainty, redundancy, ambiguity, and inconsistency are the specific characteristics differentiating IoT data. Finally, in terms of data interoperability, semantics and incompleteness makes IoT data different [32].

B. Big Data Generation and Acquisition

There are different types of sources of big data, such as enterprise data (production data, sales data, financial data, etc.), IoT data (sensors data), bio-medical data (human gene, biomeasurement data, etc.), and other sources of data (astronomy, etc.). The IoT is a significant source for big data, which also has specific features in terms of large scale (e.g., surveillance video, location, historical sensors data), heterogeneity (different types of sensors and devices), strong space and time correlation (because of analysis and inference), noise, and high-volume low-efficiency (small portion of big data is significant for analysis, e.g., traffic accident part in the whole data). Subsequent to retrieval of raw data, big data systems include a second phase, big data acquisition, which involves big data collection (logs, sensing, etc.), big data transmission, and big data preprocessing (integration, cleaning, and redundancy elimination) [37].

Another important issue in IoT data acquisition is streaming. Data from IoT sensors and devices could be coming in real time and in continuous mode. IoT streaming also requires real time data processing. Aggregation, join operations, continuous queries, and top-k monitoring are some of the significant queries to be considered in IoT and sensor systems. Aggregate queries enable the reduction of sensor power consumption with less querying on sensors. Join queries are helpful in obtaining multiple sensor data simultaneously. Stream mining provides methods for clustering, classification, outlier and anomaly detection, and frequent itemset mining in IoT stream data [32]. Jeffery et al. [184] proposed statistical smoothing for unreliable RFID (SMURF) data to clean and smooth filter RFID stream data. However, SMURF does not solve the cross-reads (false

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2773600, IEEE Internet of Things Journal

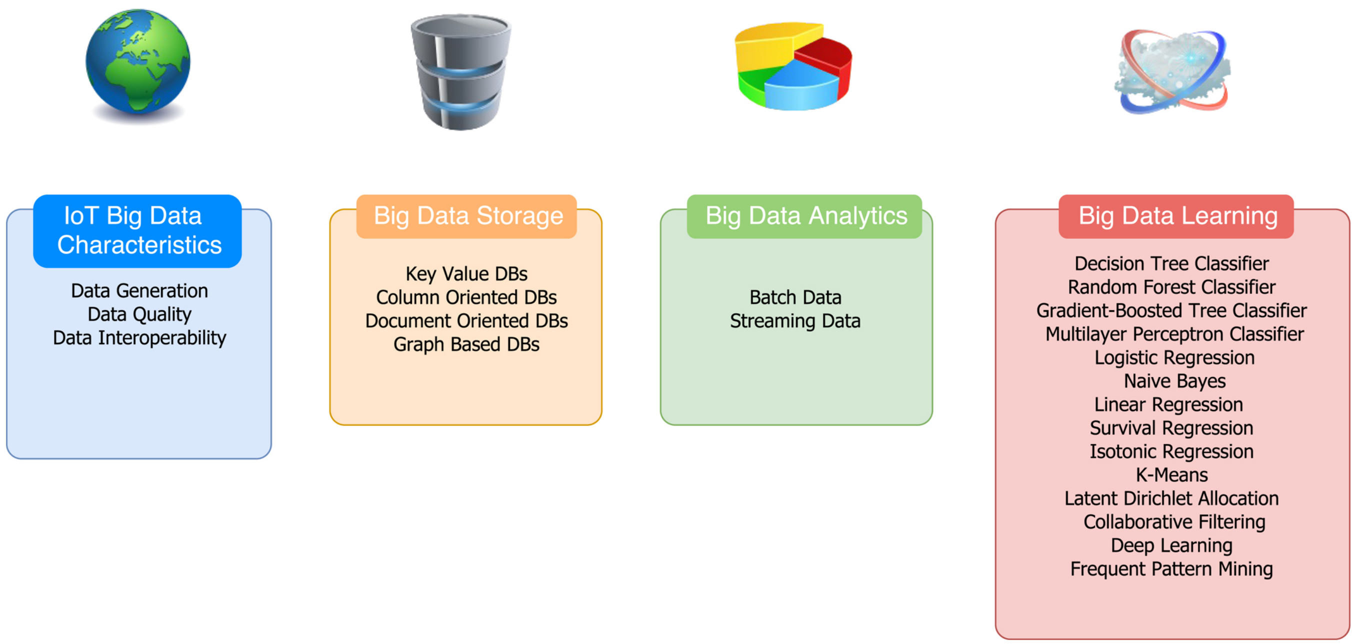

Fig. 5. Big Data in IoT

read) in the RFID stream data. Liao et al. [185] proposed KLEAP, which provided a solution for cleaning cross-reads in RFID data using a density-based method.

C. Big Data Storage

Big data storage methods are different from traditional data storage methods due to large storage requirements, managing, and analysis problems. Traditional massive data storage systems can be classified as direct attached storage (DAS), network attached storage (NAS), and storage area network (SAN). They are not suitable for big data storage. In addition to this, distributed storage systems were proposed to store massive amounts of data. However, they have critical factors that should be considered such as [37] [186]:

- Consistency: Multiple distributed storage systems should work consistently. There are also multiple copies of the same data to prevent server failure.
- Availability: More servers in the system could cause more problems and failures. All servers working without problems is the desire of user.
- Partition Tolerance: Multiple servers in the system are connected to each other via the network. If there is a failure on the network, the entire system could not work properly. All distributed systems should have tolerance for link, node, and network problems.

Distributed systems are categorized as CA (systems that have consistency and availability features), CP (systems that have consistency and partition tolerance features), and AP (systems that have availability and partition tolerance features) according to the factors that are defined [32]. These critical factors are not accomplished simultaneously [187]. Thus, storage mechanisms for big data have been developed over the years. They are categorized as file systems, databases, and programming models. File systems are the basis of applications for big data storage systems. There are different file systems in big data, such as Hadoop distributed file system (HDFS) [188], Cosmos (Microsoft development) [189], Haystack (Facebook development) [190], and Google File System (GFS) [191]. Another development for big data storage is databases. Traditional relational databases do not meet the requirements of the storage of big data. NoSQL databases became popular recently for storing big data. There are four main types of NoSQL databases: key-value, column-oriented, document-oriented, and graph-based databases [37].

- Key-value Databases: Data is stored in a key-value model, such as dictionary and hash. Keys in the model are unique and values are linked to these keys. Dynamo (Amazon) [192], Voldemort (Linkedin) [193], Azure Table Storage, MemcacheDB, and Berkeley DB are examples of key-value databases. In addition, Redis, Riak, and Scalaris (Apache), Tokyo Cabinet and Tokyo Tyrant, Memcached and Memcache DB are examples of key-value storage systems [37][38].
- Column-oriented Databases: In column-oriented databases, columns are used to split and store data. BigTable (Google) [194], Cassandra (Facebook) [195], HBase[17], HyperTable[18], and C-Store[19] [196] are the examples of column-oriented databases. Column-oriented databases are read-optimized systems, whereas, traditional relational database management systems (RDBMS) are write-optimized systems. High

[17] https://hbase.apache.org
[18] http://www.hypertable.com
[19] http://db.csail.mit.edu/projects/cstore/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2773600, IEEE Internet of Things Journal

performance querying can be achieved and developed in read-optimized systems. This can be an advantage of using column-oriented systems in the IoT [32]. Tracey et al. [197] proposed an IoT framework that store sensor data into an HBase database.

- Document-oriented Databases: In document-oriented databases, data representation is more complex than key-value databases due to document complexity. In addition, the key-value model features are preserved as well. There are different document-oriented databases. Some of them are listed as follows: MongoDB (open-source data-store in binary JSON format) [198], SimpleDB (web services of Amazon) [199], CouchDB (Apache, datastore in JSON format) [200]. Preuveneers et al. [133] used CouchDB to store data from IoT sensors in project Samurai. Cecchinel et al. [201] developed an architecture of IoT sensors in the SophiaTech campus called SMARTCAMPUS. They used MongoDB as a storage of big data for the IoT.
- Graph-based Databases: In this type of database, data is represented as a graph that is used for tasks such as network analysis. There are no rows and tables in graph-based databases. OrientDB[20], Neo4J[21], and Titan[22] represent some examples of graph-based databases [38].

All big data storage systems mentioned above are related to IoT data storage and analysis. However, there are also resource-constrained devices and things in IoT systems. For the resource-constrained devices in the IoT world, storage and energy usage become crucial and critical points. Therefore, new storage systems and databases are proposed, such as SolarStore (solar-powered storage) [202], and Antelope (DB system for resource-constrained devices). Antelope is the first DBMS that provides every sensor a data-store [203]. Another storage approach is storing data in flash memory for logging, which is called amnesic storage system [204]. Every sensor node stores data, such as audio and image, which also enables compressing, querying, and efficient organizing of data.

D. Big Data Analytics

Owing to the inadequacies of the traditional parallel processing models like message passing interface (MPI) and open multi-processing (OpenMP), some parallel processing models and engines are developed for the process and analysis of big data. MapReduce is an important and popular parallel processing programming model [205]. Hadoop [206] is one of the big data processing frameworks that implements the MapReduce programming model. Other significant parallel processing models and engines for the analysis of big data are listed as follows: Dryad (general purpose distributed parallel processing engine) [207], All-Pairs (used for biometrics, bioinformatics data mining) [37], Pregel (processes large sized graphs) [208], Spark (works faster than Hadoop because of memory caching) [38], Storm (processes streaming data in real-time) [38], Flink (batch and stream processing) [38], H_2O (parallel processing engine that contains math and machine learning libraries) [209][38][37].

For analyzing streaming real-time data, there are IoT data analysis, processing, and sharing tools. Ericsson's IoT Framework [23] is one of the tools that can combine virtual streams with local ones for statistical analysis and prediction. IBM Watson IoT [24] is another commercial cloud platform for processing IoT data. Yet another IoT analysis tool is Node-Red[25] that combines IoT data, services and devices which provide data fusion in IoT data and devices [19].

E. Big Data Learning

A higher level service in IoT big data analysis is data learning. There are now several machine learning tools and frameworks that contain many learning algorithms working on big data and processing in parallel. These include: Mahout [210], Spark MLlib[26] [211], H_2O [38], SAMOA (Scalable Advanced Massive Online Analysis) [212], Flink-ML [38], Weka [213], Oryx[27], and Vowpal Wabbit [38]. Mahout contains clustering (such as k-means), classification (such as naive Bayes, hidden Markov models, multilayer perceptron, random forest, logistic regression), collaborative filtering (recommendation engines) algorithms, modeling tools, and more. Spark MLlib contains the same tools (classification and clustering algorithms) as Mahout, but, also includes a regression model that does not exist in Mahout. Mllib also contains feature extraction, optimization, and dimensionality reduction tools. H_2O differs from the other tools because it contains many tools for deep neural networks. SAMOA is a framework that supports machine learning on streaming data. Flink-ML includes machine learning algorithms and works on the Flink platform. Weka began to develop wrappers for distributed processing on Hadoop and Spark. Oryx has machine learning algorithms for clustering, classification, and collaborative filtering. Vowpal Wabbit is different from other tools and frameworks, as it is designed for fast online learning [38].

In addition to these, big data deep learning is also another important part of big data learning. Najafabadi et al. [214] surveyed deep learning applications in big data, and deep learning challenges in terms of learning from streaming data, scalability of models, high dimensionality, parallel, and distributed computing. Further, they mentioned the challenges of big data analytics as follows: "data quality and validation, data cleansing, feature engineering, high-dimensionality and data reduction, data representations and distributed data sources, data sampling, scalability of algorithms, data visualization, parallel and distributed data processing, real-time analysis and decision making, crowd-sourcing and semantic input for improved data analysis, tracing and analyzing data provenance, data discovery and integration, parallel and distributed computing, exploratory data analysis and interpretation, integrating

[20] http://orientdb.com/
[21] http://neo4j.com/
[22] http://thinkaurelius.github.io/titan/
[23] https://github.com/EricssonResearch/iot-framework-engine
[24] http://www.ibm.com/internet-of-things/
[25] http://nodered.org/
[26] http://spark.apache.org/mllib/
[27] https://github.com/cloudera/oryx

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2773600, IEEE Internet of Things Journal

heterogeneous data, and developing new models for massive data computation" [214]. Moreover, Chen et al. [215] summarized big data deep learning (large-scale deep belief network, large-scale CNN) and challenges with regards to deep learning for high volumes of data, high variety of data, and high velocity of data.

F. Big Data Security

Security of big data is an important topic in big data studies. Privacy and security of big data topics became popular with the emergence of cloud computing, social networks, and analytics engines. In the IoT research area, big data security is also important due to the usage of storage for the IoT sensors and devices. However, in the literature, there are not sufficient studies covering the IoT-related big data security issues specifically. Big data security problems are studied and surveyed in the literature. These studies are also related to IoT big data security issues as well. Cuzzocrea [216] surveyed big data security issues and challanges under the following subtitles: security issues of big outsourced databases, privacy preserving big data analytics, big data exchange, privacy preserving big graph analysis and mining, and querying cloud enabled DBMS. Security issues are summarized and analyzed by the viewpoint of IoT usage scenarios as follows:

- Outsourced databases can create security problems in general for big data resources.
- Big data analytics cause privacy preserving problems due to the deep analysis of data.
- Data exchange between databases and IoT devices is extremely important due to security issues.
- Querying and storing encrypted data from distributed big data stores is a solution for big data systems to prevent privacy and security breaches.

G. IoT and Big Data Recent Advancements

The IoT is among the most influential topics, which can directly affect human daily life activities. Hence, many IoT solutions have been proposed and developed. Currently, leading cloud companies are developing and presenting their customers with machine learning algorithms to analyze IoT big data. We discussed and analyzed four leading IoT cloud services, namely Google Cloud[28], Amazon Web Services (AWS)[29], Microsoft Azure[30], IBM Watson[31]. All four IoT cloud services have the following features: Device management, data management, real-time streaming, big data analytics, and learning ability. Their machine learning models have also similar general purpose machine learning algorithms (classification, regression, and clustering algorithms) for anomaly detection, sensor failure, and analysis of IoT data. All these services implement machine learning algorithms on their own frameworks. Further, Google Cloud and IBM Watson use Spark ML to implement a learning structure on IoT big data. In addition, Google cloud uses Tensorflow to analyze IoT data in a deep learning structure. They also support big data storage and analytics. However, the infrastructures vary among different services. Google Cloud uses BigTable, BigQuery Streaming for storage and streaming. AWS uses DynamoDB for IoT data storage. Microsoft Azure uses a Hadoop environment for storage and analytics. IBM Watson uses CouchDb for storage.

In addition to these, open source and academic big data analytics and learning tools are proposed and developed. These are mentioned in the Big Data Learning subsection. Apache Spark MLlib[32] is the most recently developed IoT big data analytics development environment. IoT data analysis applications can be implemented with an MLlib library that consists of classification (decision tree, random forest, gradient-boosted tree, multilayer perceptron, logistic regression, and naive Bayes classifiers), regression (linear, generalized linear, regression tree, random forest, gradient-boosted tree, survival, and isotonic regression), clustering (k-means, latent Dirichlet allocation, bisecting k-means, Gaussian mixture model), collaborative filtering, model selection, and feature extraction feature algorithms. Alsheikh et al. [149] proposed mobile big data analytics using Apache Spark and deep learning. They implemented a deep learning model using Spark to analyze human activity recognition. Moreover, deep learning implementation can be achieved with Spark and other deep learning libraries and tools. These are listed as follows: SparkNet[33], CaffeOnSpark[34], Sparkling Water[35], Deeplearning4J[36] and TensorFlow on Spark[37]. SparkNet and CaffeOnSpark are projects that can run with Apache Spark and Caffe, and provides the implementation of deep learning. SparklingWater integrates H_2O with Apache Spark. Deeplearning4J is a java-based library that provides the implementation of neural and deep learning networks running with Apache Spark and Hadoop. TensorFlow on Spark can handle distributed deep learning implementation with Apache Spark and TensorFlow.

ICN provides a new perspective of implementing IoT applications. Amadeo et al. [54] mentioned that ICN will be used for a vehicular communication application. In their survey paper, they analyzed and discussed the advantages of the usage of ICN in vehicular IoT applications. The motivation of implementing a vehicular IoT application with ICN is due to the fact that vehicular communication applications are information-oriented, disregarding the producer identity and focusing on content, location, and time interval. In the future, ICN usage for implementing IoT applications will also contribute to big data-IoT studies for reducing the network load by discouraging request and packet duplication, and higher scalability. ICN could manipulate on-the-fly data using filtering and aggregating functions that reduce network resource usage and provide data retrieval scalability [54].

[28]https://cloud.google.com/solutions/iot/
[29]https://aws.amazon.com/iot-platform/
[30]https://www.microsoft.com/en-us/cloud-platform/internet-of-things-azure-iot-suite
[31]http://www.ibm.com/internet-of-things/
[32]http://spark.apache.org/docs/latest/ml-guide.html
[33]https://github.com/amplab/SparkNet
[34]https://github.com/yahoo/CaffeOnSpark
[35]https://databricks.com/blog/2014/06/30/sparkling-water-h20-spark.html
[36]https://deeplearning4j.org/
[37]https://github.com/adatao/tensorspark

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2017.2773600, IEEE Internet of Things Journal

VII. Open Issues and Future Directions

It is clear that there will be many more developments with massive deployments of the IoT in the coming years, and further research and development will be needed in the future. Below, we try to summarize some of the open issues in the area.

A. Enhancing IoT Standards

In recent years, there have been more IoT standardization efforts with the support of companies and research groups. There is still a lack of standardization in machine learning implementation and big data processing structures. As seen in the case of any emerging technology, it might take some time before a globally accepted industry standard is adopted. During this standardization process, different vendors and companies might try to implement their systems through several attempts. Eventually, the driving technology and underlying economics might be the ultimate decision maker in this adaptation process. Owing to these uncertainties, IoT hardware and software developers might be careful in their implementation strategies. However, this can also be viewed as an opportunity for technological enhancements for the overall industry.

B. Privacy and Security Issues in IoT World

There is a growing concern about security and privacy issues within the IoT networks that require distributed access, high-rate streaming data flow, autonomous decision making capabilities, etc. With the increasing complexities of such networks, it becomes much more difficult to maintain the overall security and privacy. Cyber security has become one of the most important areas due to the aforementioned problems in the IoT world. Increasing security precautions might degrade the data processing performance of the IoT/big data analytics process. However, a security breach in such a system can have devastating results, because these systems are currently integrated into several crucial industrial and/or government applications, such as transportation systems, water/sewage systems, healthcare, etc. Thus, a trade-off between security and performance might not be possible. As a result, maintaining the highest possible security level for such systems and, at the same time, maintaining the overall system performance to address real-time data processing necessities are major challenges. This is probably one of the major open issues of IoT/big data research. There are also security and privacy problems in IoT devices and frameworks under development.

C. Using Semantics in IoT World

The semantic web is used in the lower layers of IoT development, such as SSN ontology, for describing sensors and sensor data. As the heterogeneity and size of the IoT world is expanding, interoperability between devices, frameworks, and systems will become more difficult. OneM2M standard tries to overcome this, using semantic technologies for abstraction and interoperability. We will see more efforts in this direction [217].

D. Developing Learning Systems for IoT

In recent years, many learning systems and solutions have been developed with context awareness features. They are mostly designed as rule-based, logic-based, and ontology-based solutions, and developed using supervised, unsupervised, and reinforcement algorithms. They can be improved with mixed or hybrid methods, such as rule and ensemble learning algorithms for better performance. Neural networks can be used more extensively with the rapidly growing IoT big data, and solutions can be enhanced with better reasoning capabilities. Learning frameworks and systems are chiefly designed for sensor networks with limited data production and usage. New learning frameworks for larger IoT deployments with much higher data traffic can be designed and developed with new big data analytics and reasoning solutions. Moreover, unsupervised learning and sensor fusion techniques can be improved to process and analyze IoT data.

E. Implementing and Developing Deep Learning Techniques on IoT Big Data

In the last few years, deep learning has emerged as a revolutionary technique to provide robust solutions in classification and/or prediction problems where traditional machine learning models are failing. One basic characteristic of deep learning is to use the low-level features (or even raw data itself) and transform them into meaningful, high level features within the model via applying unsupervised and supervised learning through cascaded layers. As a result, better recognition is achieved in the final stages of the model. Successful implementation of deep learning generally requires huge data sets, where the model learns the hidden high-level features from data. This fits well with the big data and/or IoT concept. As a result, the number of research and application papers combining deep learning and the IoT began appearing in literature. Even though interest in big data and the IoT has been quite visible in recent years, the integration of deep learning with these fields is still an ongoing process; hence, the number of research papers combining these areas is limited. In addition, the combination of different deep learning models can be improved (RNN-Reinforcement, RNN-LSTM, LSTM-Reinforcement, etc.).

F. Developing Big Data Solutions for IoT

Big data solutions have emerged with Internet initiatives and solutions. These big data solutions may not meet the requirements of IoT data, in terms of analysis, storage, acquisition, and learning. Cloud computing solutions can be combined with the available big data solutions for the IoT. Then, big data learning and analytics solutions for the IoT can be developed for better reasoning performance [217], [218].

G. Increasing Autonomy and Implementing Self Organized IoT Structures

IoT applications can be extended and developed with fully automated M2M communication, automated reasoning, and learning systems. Without supplying information and notification to humans for interaction and decision, devices can be able to perceive, learn, interact, and decide, using IoT systems and solutions. Digital personal assistants that manage all surrounding devices, and communicate with other personal assistance solutions can be further developed. Distributed knowledge of IoT structures can be developed and improved. Smart cities that manage all IoT devices and communicate with other smart cities, autonomous vehicles, smart road systems, smart grid and energy production systems, and a smarter planet, can be future projects and initiatives using IoT solutions, which are capable of perceiving, learning, interacting, and deciding. The realization of this vision, autonomous and self-organizing IoT infrastructures, will be enabled by the ongoing research in the areas of semantics for interoperability and intelligent cognition, and learning from big data.

VIII. Conclusion

In this survey, we covered IoT studies from both historical and conceptual perspectives, as the field has evolved into a number of different dimensions. Related fields, such as ubiquitous computing, pervasive computing, AmI, WSN and its variations, have evolved with time and brought us today to what is called the IoT (section III). As the field has evolved over time, context awareness has become an essential part of the IoT. This is because it relates to understanding the changes in the environment, and provides an opportunity to act and respond accordingly. This was the first step towards intelligence in the IoT (section IV). Understanding context and acting accordingly, requires learning and many machine learning techniques were developed and adapted in the field (section V). As the number of sensors and devices increases and continues to increase at an unprecedented rate, data flowing from the IoT has become a major issue. Collecting, managing, processing, and analyzing data requires new methods and techniques as features of data in volume, variety, and veracity dimensions are completely different from traditional data. This is now called the “big data” problem, and the IoT data has suddenly become the IoT big data. As such, this problem requires new approaches. We reviewed some existing and upcoming solutions, and methods in IoT big data, and big data in general, in section VI. Finally, in section VII, we also presented the open issues that need to be addressed in the near future.

References

[1] K. Ashton. “That Internet of Things thing”. In: *RFID journal* 22.7 (2009), pp. 97–114.

[2] Charith Perera et al. “Semantic-Driven Configuration of Internet of Things Middleware”. In: *9th International Conference on Semantics, Knowledge and Grids*. IEEE, 2013, pp. 66–73.

[3] Luigi Atzori, Antonio Iera, and Giacomo Morabito. “The Internet of Things: A survey”. In: *Computer Networks* 54.15 (2010), pp. 2787–2805.

[4] Jayavardhana Gubbi et al. “Internet of Things (IoT): A vision, architectural elements, and future directions”. In: *Future Generation Computer Systems* 29.7 (2013), pp. 1645–1660.

[5] Rich; John Seely Brown Weiser, Mark; Gold. “The origins of ubiquitous computing research at PARC in the late 1980s”. In: *IBM Systems Journal* 38.4 (1999), pp. 693–696.

[6] Vlad Stirbu. “Towards a RESTful Plug and Play Experience in the Web of Things”. In: *IEEE International Conference on Semantic Computing*. IEEE, 2008, pp. 512–517.

[7] Mihai Vlad Trifa. “Building Blocks for a Participatory Web of Things: Devices, Infrastructures, and Programming Frameworks”. PhD thesis. 2011, p. 190.

[8] Dennis Pfisterer et al. “SPITFIRE: toward a semantic web of things”. In: *IEEE Communications Magazine* 49.11 (2011), pp. 40–48.

[9] Amelie Gyrard. “Designing cross-domain semantic Web of things applications”. PhD thesis. 2015.

[10] B.N. Schilit and M.M. Theimer. “Disseminating active map information to mobile hosts”. In: *IEEE Network* 8.5 (1994), pp. 22–32.

[11] Arkady Zaslavsky, Charith Perera, and Dimitrios Georgakopoulos. “Sensing as a Service and Big Data”. In: *Proceedings of the International Conference on Advances in Cloud Computing*. 2012, pp. 21–29.

[12] Phil. Wiley and Simon. *Too Big to Ignore : The Business Case for Big Data*. Wiley, 2013.

[13] Li Da Xu, Wu He, and Shancang Li. “Internet of Things in Industries: A Survey”. In: *IEEE Transactions on Industrial Informatics* 10.4 (2014), pp. 2233–2243.

[14] Daniele Miorandi et al. “Internet of things: Vision, applications and research challenges”. In: *Ad Hoc Networks* 10.7 (2012), pp. 1497–1516.

[15] John A. Stankovic. “Research Directions for the Internet of Things”. In: *IEEE Internet of Things Journal* 1.1 (2014), pp. 3–9.

[16] Eleonora Borgia. “The Internet of Things vision: Key features, applications and open issues”. In: *Computer Communications* 54.12 (2014), pp. 1–31.

[17] Andrew Whitmore, Anurag Agarwal, and Li Da Xu. “The Internet of Things A survey of topics and trends”. In: *Information Systems Frontiers* 17.2 (2015), pp. 261–274.

[18] Dhananjay Singh, Gaurav Tripathi, and Antonio J. Jara. “A survey of Internet-of-Things: Future vision, architecture, challenges and services”. In: *2014 IEEE World Forum on Internet of Things (WF-IoT)*. IEEE, 2014, pp. 287–292.

[19] Julien Mineraud et al. “A gap analysis of Internet-of-Things platforms”. In: *Computer Communications* 89-90.9 (2015), pp. 5–16. ISSN: 01403664.

[20] Charith Perera, Chi Harold Liu, and Srimal Jayawardena. “The Emerging Internet of Things Marketplace From an Industrial Perspective: A Survey”. In: *IEEE Transactions on Emerging Topics in Computing* 3.4 (2015), pp. 585–598.

[21] Charith Perera et al. “Context Aware Computing for The Internet of Things: A Survey”. In: *IEEE Communications Surveys & Tutorials* 16.1 (2014), pp. 414–454.