Journal of King Saud University – Computer and Information Sciences 34 (2022) 3569–3590

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Internet of things and data mining: An application oriented survey

Priyank Sunhare [a,b], Rameez R. Chowdhary [c], Manju K. Chattopadhyay [b,c,*]

[a] Department of Electronics and Telecommunication Engineering, Government Polytechnic College, Dewas 455001, India
[b] Department of Electronics and Telecommunication Engineering, Institute of Engineering and Technology, Devi Ahilya University, Indore, India-452017
[c] School of Electronics, Devi Ahilya University, Khandwa Road, Indore, M.P. 452017, India

ARTICLE INFO

Article history:
Received 17 February 2020
Revised 20 June 2020
Accepted 5 July 2020
Available online 9 July 2020

Keywords:
Internet of things
Data mining
Big data
Smart home
Ambient assistant living
Smart healthcare
Smart grid
Industrial IoT
Smart manufacturing
Smart agriculture
Smart transportation

ABSTRACT

Advancement in the fields of electronic communication, data processing, and internet technologies enable easy access to and interaction with a variety of physical devices throughout the globe. Our whole world is enveloped by a blanket of innumerable smart devices equipped with the sensors and actuators. Extensive research on the Internet of things (IoT) with cloud technologies, make it possible to accumulate tremendous data created from this heterogeneous environment and transform it into precious knowledge by utilizing data mining technologies. Furthermore, this generated knowledge will play a key role in intelligent decision making, system performance boosting, and optimum management of resources and services. With this background, this paper presents a systematic and detailed review of various data mining techniques employed in the large and small scale IoT applications to formulate an intelligent environment. It also presents an overview of cloud-assisted IoT Big data mining system to better understand the importance of data mining for an IoT environment.

Contents

* Corresponding author.
E-mail address: mkorwal@yahoo.com (M.K. Chattopadhyay).
Peer review under responsibility of King Saud University.

ELSEVIER Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2020.07.002
1319-1578/

1. Introduction

MANY research communities and industries all over the globe are vigorously pursuing research and contributing a lot in lifeless objects (things) to make them live and work smartly. The global giants including Apple, Microsoft, Google, IBM, Cisco, Siemens, Huawei and Research communities including Internet of Things (IoT), Mobile Computing (MC), Wireless Sensor Networks (WSN), Machine to Machine communication (M2M), Pervasive Computing (PC), Cyber-Physical System (CPS) etc., around the globe, are working persistently for the formation of new Concepts and Standards to create Smart environment (Stankovic, 2014; Kravchenko et al., 2017; Miorandi et al., 2012).

The basic idea of Internet of Things is that the things/ objects can be connected to the internet. They should have a unique identity, should be identified automatically, and should communicate with each other and the humans. They should make decision by themselves or follow the human commands (Tsai et al., 2014). So it's not wrong if we say, in IoT, internet can be considered as a global platform which powers machines and smart things to communicate, compute, make decisions and coordinate with the humans globally(Miorandi et al., 2012; Bandyopadhyay and Sen, 2011).

The question now arises that why Research communities and Industries around the globe have a sudden interest in the Internet of Things? Researchers around the world hold a view that the cities and world itself will be overlaid with the sensing and actuation devices in the next 5–10 Years. The density of sensing and actuating devices will increase many more times than the population of the world. Ericsson and Cisco predicted that "50 billion of small embedded sensors and actuators will be connected to the internet by the end of year 2020 and Internet of Things will create 14.4 Trillion dollars of value at stake for industries in the next decade" (Google, 2017). So we can say, an ultra large number of connected heterogeneous devices will form the Internet of Things environment (Stankovic, 2014; Kravchenko et al., 2017; Miorandi et al., 2012; Uusitola, 2006; Ortiz et al., 2014: Bijarbooneh et al., 2006; Yue et al., 2014).

The heterogeneous smart IoT environments generate enormous amount of data. This data is initially a raw data that needs to be processed to be of any use. This raw data either belongs to the Infrastructure i.e. it is an infrastructure centric data containing devices and network related information; or the data belongs to the environmental parameters of IoT constructed environment (Tsai et al., 2014) i.e. it contains sensor – recorded outputs, actuator actions etc. Both types of data here are inherently heterogeneous in nature. If processed well with an appropriate data mining algorithm, one can dig out valuable knowledge from it (Chen et al., 2015).

A challenging part of any IoT enabled smart environment is to select or synthesize the most appropriate data mining algorithm. Such an algorithm should produce valuable analytics, predict future events precisely and manage the network and services efficiently within all constrains. To understand this in depth, observe Fig. 1. The figure depicts several IoT applications ranging from small to large scale. Each block represents an IoT application that

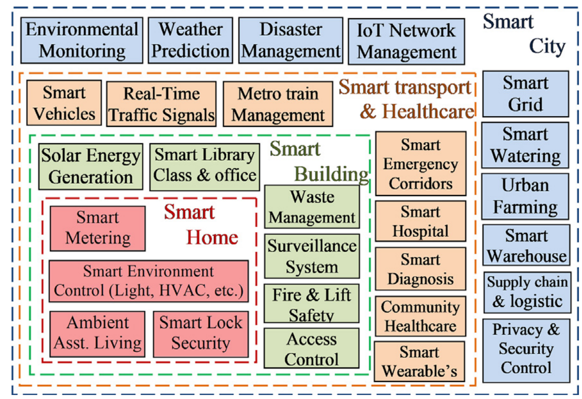

Fig. 1. IoT Applications: from Small scale to Large scale.

performs a particular task. For e.g., Ambient Assistant Living (AAL) (Youngblood and Cook, 2007; Samarah et al., 2017; Joergschmalenstroeer, 2010) where we provide technologically supported smart environment, such as healthcare assistant, in home itself. AAL utilizes number of smart sensors, actuators, wearable's, smart gadgets, CCTV cameras and communication devices with inter-network connectivity. The devices in AAL environment have different capabilities and limitations. They are connected to each other via a variety of communication mediums (like Bluetooth, Wi-Fi etc.). The raw data produced from these devices also differ in characteristics. Converting this raw data into semantically meaningful corresponding activities so as to identify the user's present state, requires data mining. Therefore, when we process this new sort of data with the traditional data mining algorithms, it does not provide an accurate insight. And, our system may not establish an intelligent and responsive environment. Discussion of this single IoT application highlights the diversity of devices and their respective data.

If move diagonally up from left side lower corner in Fig. 1, we find IoT applications that comprise of several other small scale applications. For example, the smart home envelopes four applications *viz.* the smart metering, the smart environment control, the AAL, and the security. A multistoried smart building may have in it: multiple smart apartments, smart offices, a library, classrooms, a waste management system, a surveillance system, a smart lift and a fire safety and access control management (Pacheco et al., 2019). So a smart building can be defined as a fusion of technologies which give thinking capability to the building so as to provide a convenient, comfortable, efficient and healthy environment to its habitants.

Smart transport and healthcare Application has a number of smart buildings (residential, hospital, commercial, schools and colleges) along with the other applications such as smart vehicles, real-time traffic control, metro-train sevices, emergency corridors, disease diagnostics and healthcare being operated together for improved quality of living. Last one in Fig. 1 is the smart city application that envelops all: smart home, smart building, smart transport, smart healthcare and the other essential applications *viz.*

environmental monitoring, weather prediction, disaster management, IoT network management, smart grid, urban farming, smart watering, smart warehouse, supply chain and logistics, and importantly, the device, network privacy and security control management.

Smart objects like sensors, actuators and embedded devices, in this complex environment of IoT applications, produce and consume huge amount of data. Hence, the knowledge extraction (Data Mining) mechanism can be considered as the heart of the complete system. Better information extraction provides efficient value-added services to the end users.

Huge numbers of devices, with basic to advanced features, are being connected in small to large scale application environments. In such a scenario, it is possible that a single device may serve more than one applications. Such situations require a centralized middleware so as to ease the development process, create useful analytics, provide privacy, security and trust mechanism as well asto support interoperability within diverse applications and services (Razzaque et al., 2016). Consequently, we predicted astronomically huge heterogeneous network of IoT devices with data mining algorithms to subsist as a seamless fabric, covering and synthesizing the intelligent environment (Miorandi, et al., 2012).

1.1. Characteristics and challenges

There are several characteristics of IoT infrastructure and IoT data mining algorithms/technologies. These include unique identification of each device, ultra large scale network of things, devices and network level heterogeneity, ubiquitous integration, interaction and interoperability, robust data and devices management, dynamic entry and exiting of devices to the network at very fast rate, service oriented computing, Real-Time/Resource-constrained devices and privacy, security and trust management (Razzaque et al., 2016; Xu and Helal, 2016; Kantarci and Mouftah, 2014; Stojmenovic, 2014). All these characteristics are converting research and development of the diverse applications and services into newer challenges:

Unique Identification of each device

- *IoT infrastructure:* Identification is the basic need to establish communication between smart objects. Generation and management of unique Id's globally for trillions of devices is a very challenging task. From internet perspective, IPV6 may be helpful but what about the devices communicating via alternate mediums like Bluetooth, NFC, Zigbee and many others? Can RFIDs, QR codes or similar IDs be the solution? Or should we adopt name centric network architecture instead of host centric? Robust naming protocols are lacking and are required globally.
- *IoT data mining:* From IoT data mining perspective too, unique identification play a very crucial role. Better understanding of data related to infrastructure (like unique id, device features etc.) can result in better actuation control from derived knowledge. Acquiring, storing and managing unique id of trillions of devices and related features is another big challenge.

Ultra large scale network of things

- *IoT infrastructure:* The blanket of IoT infrastructure is fabricated with ultra large scale of sensors, actuators and embedded devices where they serve human needs intelligently. To develop a mechanism that supports globally unique identification, authenticate access during operation, maintenance, and protective utilization on such an ultra large scale are creating several challenges. There are even more issues that need to be addressed while dealing with massive scaling are the longitudinal studies for the deployment of smart devices, environment changes over the time, and self- management/automation with maximum service utilization of resource constraint devices (Stankovic, 2014; Kravchenko et al., 2017; Miorandi et al., 2012).
- *IoT data mining:* The huge network of devices generates a new type of data known as the IoT big data. The biggest challenge in today's data mining world comes with several issues like data storage, management, privacy, security, and processing limitations such as real-time/streaming data. Instead of gathering all data in the servers, the data pre-processing techniques like the data filtering, dimension reduction, feature selection, pattern reduction play crucial role.

Devices and network level heterogeneity

- *IoT infrastructure:* IoT is a network of versatile devices, where the devices from very few features, basic computing, low memory, low power and energy consumption (like RFID, QR codes, nano-sensors and actuators, MEMS etc.) to the devices having advanced features, extraordinary processing capabilities and larger memories are internetworked and slog together for a specific or multiple purposes. In many applications, sensor nodes, devices and machines are interconnected using limited communication mode (for e.g. Bluetooth and NFC) that make network cluster, and any one higher end device of that cluster, connect to a global network. Therefore, IoT is a highly dynamic and radically distributed network of things. Synthesizing the robust architectures and protocols support and accommodate heterogeneity. How to provide better management at all level of technologies, services and applications, is a new challenge for the future research. Concurrent and massive access communication between machines over radio access networks (i.e. 5G networks) may cause subsequent performance degradation, including service unavailability, intolerable delay, and packet loss (Oh et al., 2015). Ubiquitous parallel computing and storage with improved existing techniques and algorithms seem to be the feasible solutions here.
- *IoT data mining:* Compared to the data mining results of traditional huge data set, heterogeneous IoT big data mining maximize overall potential with finer knowledge and insight of targeted application area. Thais creates series of new opportunity and challenges. It an obligation not option, that one has to deal with the structured, the semi-structured and the unstructured data all together. Applying data mining in sensor networking and device management yield superior results for resource constraint environment (e.g. LEACH (Ankerst et al., 1999), DataCloud (Yue et al., 2014)). But yet, it is limited to only particular applications and hence is one of the biggest challenges. Heterogeneous network here not only access or extract information from larger scale data, but it also needs to deal with the dynamic, uncertain and incomplete data.

Ubiquitous integration, interaction and interoperability

- *IoT infrastructure:* The ubiquitous embedded sensors/actuators with highly versatile characteristics and protocols participate, and, are shared among many applications in an IoT environment. The distinctly critical part is to make automated device integration at immensely large scale, and to ensure interoperability among them. Research is needed to incorporate a centralized middleware core that can resolve dependencies across the applications; support *trans*-coding between the assorted protocols; perform efficient interaction among the devices; and dynamically create and provide innovative services.
- *IoT data mining:* Human to human, human to machine and machine to machine are the three type of interactions that normally happen in an IoT network. The data from the same device may have different meaning according to the type of interaction, application and service. Secondly, the same type of data from various devices, applications and services may also not pose exactly the same meaning. And there is not a single device

or application, it's a huge network. Creating the data mining algorithm for such a huge environment is highly critical. Algorithm should have the capability to extract knowledge as per the specific service and application needs, so that the keen interaction and interoperability can be fabricated.

Service oriented intelligence

- *IoT infrastructure:* Always ON responsive services are the inherent property of IoT environment. It supports every day user needs. There are cases where number of devices are either mobile or battery powered that may leave and rejoin network many times or they may even randomly join network for a specific purpose. In this ubiquitous ever-changing environment, IoT infrastructure also sometimes demand ad-hoc applications and services which can be composes, executed and demolished in runtime irrespective of whether they have been considered while architecting the system.
- *IoT data mining:* The IoT applications environment equipped and deployed with a lot of IoT sensors, utilizes tremendous services *viz.* real-time interaction among connected devices, system management, ad-hoc applications management, privacy and security management and many more. The management of flexible, dynamic and open ad-hoc IoT services becomes possible due to highly precise data mining algorithm. Incorporating dynamic knowledge synthesizing capability for a randomly originated ad-hoc services is one of the most challenging part of an IoT data mining algorithm.

Privacy, security and trust management

- *IoT infrastructure:* Privacy, security and trust should be integrated at all levels in a system. Many IoT applications like healthcare, emergency systems, Physical access control and many more are serving human in critical conditions. If someone hacks the devices and network, it will be a big threat for many lives. Not only the security but maintaining privacy and trust is equally important. Let us take example of a smart phone; that has number of apps installed including banking, social media, healthcare, entertainment along with other personal data. Wearable sensor, healthcare gadgets, smart appliances and other devices are connected to this phone via variety of mediums. The phone is maintaining not only one's privacy and security, but is also sharing many important details with the network connected devices for uninterrupted smart environment services. Now, if any of the devices has leakage, the person's data will no more remain private (Pacheco et al., 2019; Qiu et al., 2019). Developing an interface that can detect attacks and enforce privacy, trust and security in this diverse environment is a challenging task.
- *IoT data mining:* As the privacy and security mechanisms are designed for a certain amount of data with specific characteristic. IoT systems with huge data also poses highly divers features set. It contains larger features set of our private data. The data mining algorithms are applied over input data that can extract several useful and personal information too. This can enable the attacker to weaponize the data. Ad-hoc services in web based systems are one of the major threats for privacy and security applications which can be resolved using deep learning for data mining (Pacheco et al., 2019). Consecutively, one needs to address this issue at each level in a distributed manner with adequate mining and management solutions.

Therefore with the specified characteristics, IoT infrastructure and Machine Learning creates robust intelligence i.e. the capability to perceive, reason, decide, perform actions, learn and interact (Youngblood and Cook, 2007). Research communities and industries are persistently diligent in making this possible in the real world. Often, researchers around the world find some good implementable ideas, but these ideas are tested on small/limited problems or simulated on virtual platforms which do not scale appropriately for the larger and complex real-world problem (Yue et al., 2014; Joergschmalenstroeer, 2010; Rashidi et al., 2011; Saives et al., 2015; Zdravevski et al., 2017; Virone et al., 2008). Also, the research on IoT is highly fragmented and so thinking of global solution may not be achievable. We require some standards that should be stated globally to lead researches in particular direction.

1.2. Literature review and contribution

Numerous good surveys over IoT and its data have been presented from different perspectives. Stankovic (2014) highlighted the vision and characteristics of IoT from a global perspective with strong and a very informative discussion over eight research areas. It also suggested an architectural approach for IoT to borrow from Smart-Phone world i.e. enabling App-Store like environment to ace the development, authentication, installing and uninstalling applications and services. (Miorandi et al., 2012) and (Bandyopadhyay and sen, 2011) presented the survey from the viewpoint of technologies utilized in IoT with the possible research and applications. Reference (Bandyopadhyay and sen, 2011) also presented generic five-layer architecture for IoT system design. Five layers from bottom to up include edge technology, an access gateway, Internet, Middleware and application, whereas in (Razzaque et al., 2016), the survey is dedicated to IoT middleware. It focuses on computing, interoperability and communication within the heterogeneous environment of applications and services. Dai et al. (2019), have investigated the integration of Blockchain technology into IoT Architecture. They introduced in-depth, the blockchain technologies with IoT and also presented Blockchain of things (BCoT) architecture with several benefits of 5G connectivity environment. Phuttharak and Loke (2019) imparted an extensive survey on mobile crowd sourcing research, highlighting the aspects of implementation needs during the development, architectures, and key considerations for their development. All abovementioned surveys are centered on architectural challenges of IoT infrastructure, with not much concern about the data mining algorithms.

There are several more researches that surveyed the convergence of data mining with IoT. References (Tsai et al., 2014; Chen et al., 2015; Marjani et al., 2017; Mohammadi et al., 2018) presented a strong and systematic review of data mining algorithms from IoT perspective. Considering IoT environment, Tsai et al. (2014) introduced core data mining algorithms from "data about things" and "data generated by things" perspective with a unified framework that included scan, construct and update functions. Chen et al. (2015) presented knowledge, technical and application view. A one step forward, (Marjani et al., 2017; Mohammadi et al., 2018; Nahar et al., 2019) discussed the upcoming research challenges due to a new type of big data i.e. the heterogeneous and devices engendered IoT big data. Marjani et al. (2017) investigated the Power of IoT big data analytics in IoT applications. With the discussion of IoT big data analytics, method and techniques, they also presented a cloud oriented IoT big data architecture. (Marjani et al., 2017; Mohammadi et al., 2018) carried out survey on IoT real time big data streams and provided an in-depth overview of deep learning algorithms and architectures that foster better analytics and learning. In (Mohammadi et al., 2018), authors also summarized in detail the major research attempts that leveraged deep learning and the approaches supported by Fog and cloud in an IoT application environment. Although the above surveys on the IoT and data mining are strong enough and provide anin-depth learning and utilization of data mining in IoT, they have highlighted the IoT applications part only briefly.

Because of fragmented research on IoT, the solutions developed while considering one application environment may not support other. Therefore, there are diverse survey papers presented on IoT and data mining from the perspective of multiple applications. The few are:

- M. Rashid et al. (2020) discussed and critically analyzed the existing behavioral pattern mining algorithms. They also proposed knowledge based framework for real-time stream data of numerous sensors in WSN and IoT.
- Qolomany et al. (2019) conducted a very knowledgeable survey on smart building from the viewpoints of applications, data analytics, and machine learning.
- Qi Chen et al. (2019) provided technical-oriented and application-oriented review of smart city convergence with the deep learning.
- References (Shu et al., 2018; Pacheco et al., 2019) surveyed big data mining and machine learning integrated healthcare and large scale petrochemical plant applications, respectively.
- Pacheco et al. (2019) conducted a systematic survey to explore the deployment of machine learning techniques to achieve the network traffic classification.
- As most of the knowledge extracted contains highly private data of any user, Qiu et al. (2019) attempted survey in the field of access control for search engines of IoT environment data from security perspective.

Survey papers mentioned above pursue effective and knowledge generating research. Most of the research is technology, knowledge extraction or analytics oriented. Some presented the applications view also, but were specific to a particular application *viz.* smart vehicle or smart city. As we discussed, a wide range of applications are fortified by IoT. Many large scale applications may comprise lots of small scale applications within it.

Therefore, this paper presents an application-oriented systematic and detailed review of various Data mining algorithms and their variants that are well utilized in an IoT environment. In this application-oriented survey, the major contributions of our research work are:

- The article explores the numerous IoT applications environments and identifies their potential integration with diverse data mining algorithms.
- In the introduction part, we present edinter-correlation of IoT applications for better understanding of the convergence of IoT and data mining. Here we have also highlighted the characteristics and its associated challenges.
- We propose an IoT big data mining system that provides an overview of overall complex intellective environment for applications ranging from data extraction to the processing and then service execution.
- Introduction to key data mining algorithms with their variants and their utilization for assisting several intelligent operations related to the applications like Smart Home, Ambient Assistant Living, and Smart Healthcare, Smart Grid, Industrial IoT, Smart Manufacturing, Smart Agriculture and Smart transportation.
- Lastly, we summarize the applications environment and related open research issues with suggestions for the research aspects of data mining and IoT with application perspective.

The remaining part of the paper is organized as follows. Section 2 gives an IoT big data mining system overview containing six layers *viz.* sensing and actuation, Gateways or pre-processors, Internet, decentralized data centers, decentralized data processing and control (knowledge discovery), and centralized processing and control. Section 3 discusses the key data mining algorithms with simple examples. Section 4 provides an in-depth survey of IoT Applications including Smart Home, Ambient Assistant Living, and Smart Healthcare, Smart Grid, Industrial IoT and Smart Manufacturing, Smart Agriculture and Smart transportation with data mining perspective. Section 5 summarizes all in tabular form with open research issues and finally the paper is concluded in Section 6.

2. IoT big data mining system overview

Billions of devices in smart environments can interact and communicate with devices all around and humans as well. This engenders a plethora of heterogeneous data. As very well explain in (Tsai et al., 2014), the knowledge extracted from this raw data can be categorized as the data of IoT infrastructure (i.e. unique ID, type of device, limitations, location, connectivity and mobility etc.) and the data measured by the IoT environment (i.e. measured external parameters, device to device and device to human intercommunication, data interchange and data use logs etc.). The knowledge extracted from both type of data, are equally important as the former if optimized properly can drastically improve the performance whereas the later qualitatively enhance the services of IoT infrastructure (Tsai et al., 2014). So the most paramount question that arises now is how to extract higher-level useful information from raw data. Representations of these raw data as Machine interpretable and human understandable information become the need for current IoT infrastructure (Ganz et al., 2015). For constructing knowledge from the raw data we require various data mining and knowledge discovery algorithms. As the data is heterogeneous by nature, it demands more than one processing (data mining and knowledge discovery) algorithms to work in parallel.

In this section, we present IoT big data mining system architecture. There are a number of architectures already been proposed from different IoT domain perspectives (Bandyopadhyay and Sen, 2011; Dai et al., 2019; Marjani et al., 2017; Mohammadi et al., 2018; Rashid et al., 2020). For example, Dai et al. (2019) presented Blockchain of things (BCoT) architecture where they introduced Blockchain chain composite layer between the network layer and application layer with benefits of 5G connectivity environment. Marjani et al. (2017) presented a cloud oriented IoT big data five layer architecture (from bottom to up IoT devices, network devices, gateway, cloud and data analytics) that investigated the Power of IoT big data analytics in IoT applications. By considering above references in our big data mining system architecture we undertake relationship among various applications like Smart home, Ambient Assisted living, Smart healthcare, Smart Traffic and Parking systems, Industrial IoT, Smart Agriculture as the key contribution areas of IoT to shape the Smart world. Fig. 2 shows a cloud-assisted system architectural overview for the IoT infrastructure. The lowest layer of system architecture comprises of various sensing and actuation devices. It includes sensors, actuators, a camera and several small embedded systems for home automation, healthcare, traffic, parking, automobile, Industries and Agriculture, serving various applications to perform intellectively. The raw data produced from various devices like time-series data and sequence of events detected, visual and audio data etc. was accumulated by gateway layer and pre-processed to remove noise. Repeated sequence/events use various types of Gateway processing unit's *viz.* Wi-Fi, Bluetooth or ZigBee routers with other electronic devices like smartphones and small scale embedded systems. Even the local server can be a gateway. Other than noise removal, the heterogeneous data of IoT environment requires feature extraction and data fusion and projection to be performed by the gateway layer.

The abundance of data generated from the IoT, gives birth to the new challenge - known as Big IoT data. The inherent charac-

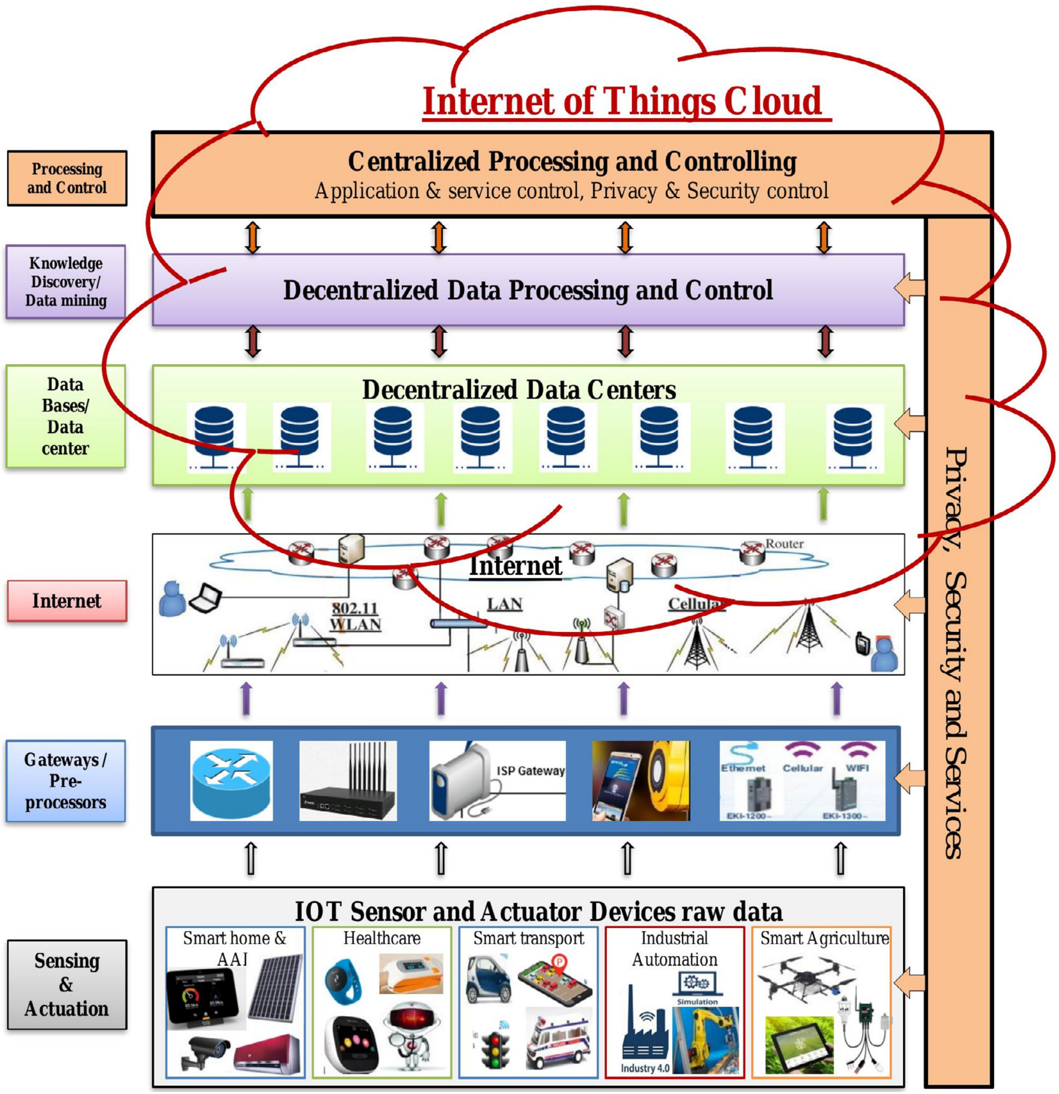

Fig. 2. Cloud assisted system architecture for Internet of Things.

teristics of such raw data are large volume, heterogeneity, velocity of generation and fast changing data (GGL, 2017). IoT Big data have time as an integral dimension i.e. it should be processed in real time or in particular short period of time otherwise after a certain deadline, the result of processing will of no use (Che et al., 2013).

After gateways, the pre-processed data is sent to decentralized data centers via internet. Decentralized processing and controlling stations then extract knowledge by putting into service the diverse data mining algorithms and the machine learning mechanism at their respective ends. This is the most important part, as the single applications are many times composed of several other applications and services. The knowledge extraction and analytics generation can vary according to the task. If the applications or the services are known beforehand, a well-defined existing mining algorithm can extract the knowledge. But in the case of ad-hoc applications or services, the system should have dynamic machine learning capabilities that can handle outliers, do modification in the model and select appropriate algorithm or a set of algorithms to extract knowledge and prepare relevant analytics.

After knowledge extraction and analytics creation, the decentralized units provide services and perform required intelligent actions in their own restricted environment. At-last, the extracted higher-level useful data of IoT infrastructure is converted into Machine interpretable and human understandable information by Centralized Processing and Controlling station in coordination with the Decentralized Data processing and control servers and, if required, can be adapted to make sensible decisions to optimize performance and the quality of services for the IoT applications and its infrastructure.

The Centralized Processing and Controlling station with the Decentralized data processing and control servers can amend the Privacy and Security. As the number of smart devices (like wearables) connect to the network through various other higher end devices (like smart phones), to establish initial communication lots of data is shared among all. Lee et al. (2018) have performed research on several smart gadgets and found that the huge private information was being shared among devices. Therefore, there can be a big threat for the privacy. Other than this, several ad-hoc services need to be served. The search engines can generate numerous queries and tasks. Not only the device data, but also our social network and healthcare data is accumulated and processed. In cloud based system architecture, the complete data is collected and processed centrally. This can invite abuse of the information collected and of the concentrated web servers (Tian et al., 2020a). The large scale enterprises also require a robust access control mechanism to manage physical access (Geepalla et al., 2013). Therefore, high security and privacy should be maintained in hybrid fashion at each level. Tian et al. (2020a,b) propose the web attack detection system that utilized distributed deep learning mechanism. Inspired by (Tian et al., 2020a) in our system architecture also we proposed a cloud based centralize management in coordination with decentralized control of privacy, security and services (Tian et al., 2020a, b; Geepalla et al., 2013; Lee et al., 2018).

Concluding this section, several research groups around the globe are vigorously pursuing research to invent smarter knowledge discovery techniques that will extract higher-level useful information from the big IoT data.

3. KeyData mining methods

The environment around us is filled with a plethora of heterogeneous data. It does seem infeasible to make this environment keenly intellective without appropriately utilizing the data mining technologies. Data mining can be a supervised, unsupervised or reinforcement learning with automation today. The computer-assisted learning grows more precisely when performed in multiple layers in a hierarchical manner. This automatic feature extraction through supervised or unsupervised learning in a hierarchical manner is known as Machine Learning (ML).

Che et al. (2013) survey for big data mining brings focus on the challenges like variety, heterogeneity, scalability, velocity, accuracy, trust, provenance, privacy crises, instructiveness and most importantly on garbage mining. The necessity of applying data mining is consequential not only for knowledge discovery but also for the Garbage elimination from the internet. Ganz et al. (2015) suggested Data Abstraction as one of the appropriate methods. They reviewed various Abstraction techniques and proposed a piece of advice to preserve only abstracted results instead of whole data in the data centers. Data mining is an integral part of knowledge discovery, as shown in Fig. 3. Data accumulated from various IoT devices are first sent to a pre-processing unit where several actions (like feature selection and extraction, noise abstraction, Normalization dimension reduction etc.) take place to mold raw data into the appropriate format for analysis. Formatted data is then sent to Data Mining unit where various data mining techniques perform their task to extract higher-level useful information. The combination of both i.e. data pre-processing and Data mining units come under a single box of DL. Further, the output of DL is evaluated and represented into machine interpretable and human understandable knowledge which is utilized further by the IoT infrastructure (See Fig. 4).

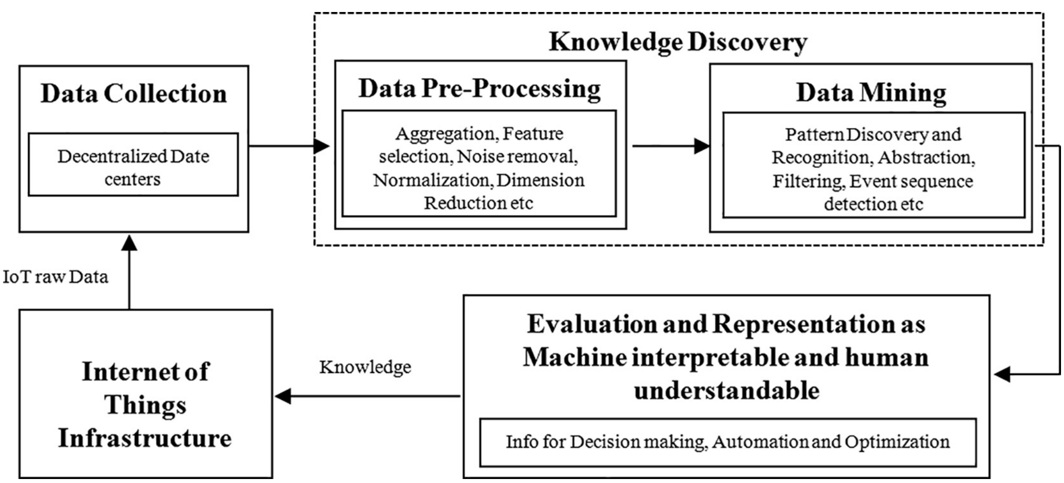

Fig. 3. Knowledge discovery overviewed.

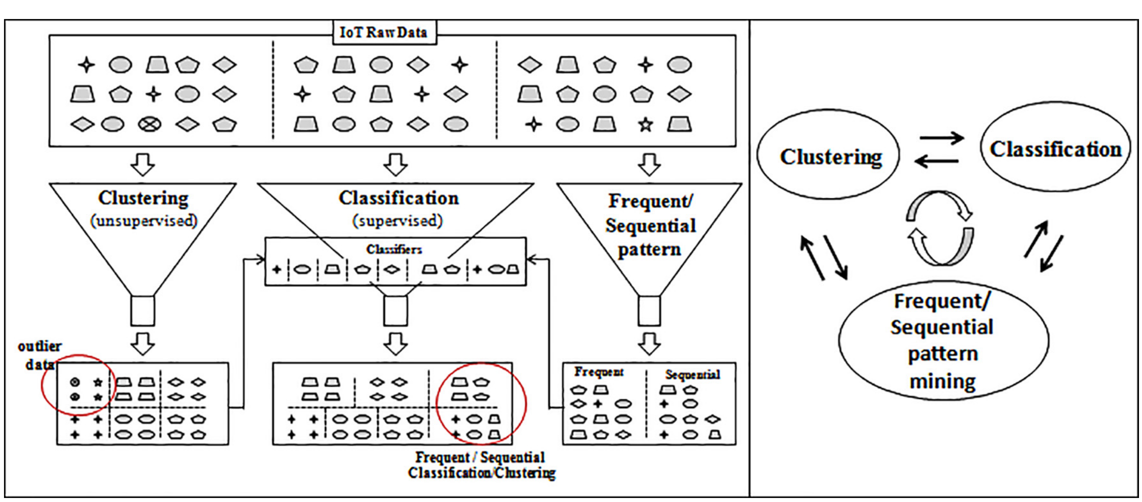

Fig. 4. Data Mining process for Internet of Things environment.

3.1. Classification

Classification is a process of assigning the objects to previously defined categories. It aims at predicting accurately the destination class for each object of data (Kesavaraj and Sukumaran, 2013). As the targeted labels are assumed to be known before processing, it is a supervised learning process (Han et al., 2007; Tan et al., 2006; Kesavaraj and Sukumaran, 2013; Liu, 2011). The prediction function (classifier) in classification requires training before being used to classify unlabeled or unknown objects/data. And so, one can use labeled or known data to train prediction function. For example in a certain medical care centre, there are data related to patients suffering from a disease having three stages as primary, moderate, critical with three specific methods to cure them as Treatment_p, Treatment_m and Treatment_c respectively. First, the classifier/prediction function is constructed from a set of rules defined by a medical researcher or by data recorded previously during the treatment. The data available is divided into two parts i.e. Training set (labeled) and testing set (unlabeled). The training set first constructs the classifier and then the test set validates it. After that, the classifier analyzes the patient data (unlabeled data) to put them into classes (Treatment_p, m or c) according to a stage identified by the constructed classifier.

Most algorithms are classified in two steps: first compute the probability of item belonging to the particular class. Second, compare it to the cutoff value and classify accordingly. Performance evaluation (Tan et al., 2006) of the classification model is defined based on the number of instances that are assigned to the right category (i.e. Accuracy) and assigned to the false category (i.e. Error rate) given as:

$$\text{Accuracy} = \frac{\text{Number of correct predictions } (\text{Tp} + \text{Fp})}{\text{Total number of predictions } (\text{Tp} + \text{Fp} + \text{Tn} + \text{Fn})} \tag{1}$$

$$\text{Error Rate} = \frac{\text{Number of wrong predictions } (\text{Tn} + \text{Fn})}{\text{Total number of predictions } (\text{Tp} + \text{Fp} + \text{Tn} + \text{Fn})} \tag{2}$$

There are also alternate Accuracy measures of classification algorithm results i.e. *precision*: defined as the probability that a randomly selected result is relevant and *Recall*: defined as the probability that a randomly selected relevant object is retrieved. This can be mathematically described as:

$$\text{Precision}(\text{Pr}) = \frac{\text{Tp}}{\text{Tp} + \text{Fp}} \tag{3}$$

$$\text{Recall}(\text{R}) = \frac{\text{Tp}}{\text{Tp} + \text{Fn}} \tag{4}$$

where Tp, Tn, Fp, Fn can be defined using below confusion matrix:

	Prediction Class = 0	Prediction Class = 0
Actual Class = 0	True Negative (Tn)	False Positive (Fp)
Actual Class = 1	False Negative (Tn)	True Positive (Tp)

On the basis of Precision and Recall, the overall classification results are described by F-score, given as:

$$\text{F} - \text{score} = \frac{2 * \text{Pr} * \text{R}}{\text{Pr} + \text{R}} \tag{5}$$

There are many classification models available to classify data into various classes depending on data characteristics and situations *viz.* Decision Tree Induction, Bayesian classification, Rule-based classification, classification by Backpropagation, Support vector machine, k-Nearest Neighbor, Deep Neural Network and Ensemble methods. A set of classifier can also deploy the fusion of various classification techniques for complex large scale IoT Application problems (Chen et al., 2015; Marjani, et al., 2017; Han et al., 2007; Tan et al., 2006; Kesavaraj and Sukumaran, 2013; Liu, 2011; Alsabti et al. 1998). Some majorly used top-rated classification methods include C4.5 (Alsabti et al. 1998), a descendant of CLS and ID3. It engenders classifier in the form of a more comprehensible set of rules. C4.5 was then superseded by C5.0 with much-improved efficiency, scalability and boosted performance by overcoming the disadvantages like large CPU Time and Memory requirement (Tsai et al., 2014). Rule-based classification, Support vector machine, Association rule analysis based classification type models of classification are highly suited for today's IoT environment. In references (Lai et al., 2013; Fleury et al., 2010; Brdiczka et al., 2007; Li et al., 2017a–c), researchers use Classification models with frequent data mining techniques like Hidden Markov Model for creating a more keenly intellective and responsive environment. The Naive Bayes, Gaussian naive Bayes, Bayesian belief network, Bayesian network, Artificial Neural Network and Ensemble methods are used over various sensors and actuators data in the applications majorly related to Biomedical, Environmental Prediction, smart building access controlling and user activity recognition, improving Sensor network Efficiency, Optimization, and Artificial Intelligence (Misgeld et al., 2016; Rad et al., 2017; Wu, 2009; Abedin et al., 2017; Rad et al., 2014; Perera and Dias, 2011).

3.2. Clustering

A cluster is represented as a group of like objects. Clustering algorithm classifies the collected objects into certain numbers of clusters where the objects in a particular cluster pose similar features. Unlike Classification, Clustering is an unsupervised learning technique (Yue et al., 2014; Tsai et al., 2014) i.e. it will not require prior knowledge to guide the partitioning process (Han et al., 2007; Tan et al., 2006). For example in certain medical care centre, it is found that a number of patients are suffering from an unknown disease. The medical researchers have the data available only about the observed symptoms and the progress made by the patient after pursuing numbers of treatments. In this kind of situation, the clustering will help through classifying the patients into a number of groups for proper treatment as per recognized symptoms and past treatment data available. The output of the clustering is represented by the set of centroids (c_n) and quality of clustering measures in Sum of the Squared Error (SSE) as given below (Han et al., 2007):

$$c_n = \frac{1}{m_n} \sum\nolimits_{x \in C_n} x \tag{6}$$

$$\text{SSE} = \sum_{n=1}^{K} \sum_{x \in C_n} \text{dist}(c_n, x)^2 \tag{7}$$

where x is an object, C_n is an n^{th} cluster, c_n is a centroid of the n^{th} cluster, m_n is the number of objects in n^{th} cluster, K is a number of clusters and *dist* is the standard Euclidean distance between two objects.

Most of the clustering methods like k-mean, K-Nearest Neighbor (Dhillon et al., 2004; Jin et al., 2006), k-medoids (Li et al., 2017a–c),Hierarchical clustering (CURE (Guha et al., 1998), SVD (Berry and Browne, 2005), ROCK (Guha et al., 1999), BIRCH (Zhang et al., 1997)), Density based clustering (DBSCAN (Ester et al., 1996), OPTICS (Ankerst et al., 1999), DENCLUE (Hinneburg and Keim, 1998)), Grid based clustering (STRING (Wang et al., 1997), WaveCluster (Sheikholeslami et al., 1998)), are designed from a single system perspective where centralized data more or less belong to certain characteristics.

With the advancement in sensor technologies, IoT and WSN formulate the user environment to be smart enough to detect user activities and act accordingly. References (Samarah et al., 2017; Rashidi et al., 2011; Brdiczka et al., 2007; Saives et al., 2015; Li et al., 2017a–c; Zdravevski et al., 2017; Virone et al., 2008) used clustering as a core technology to classify numerous features that identify an individual's daily activities and automate their tasks to increase comfort and security. In an environment such as IoT and WSN, cloud-based distributed clustering is more important than centralized clustering as the data and the devices are highly heterogeneous and hence may require different processing techniques (Younis and Fahmy, 2004; Uckelmann et al., 2011; Heinzelman et al., 2000). Saives et al. (2015) proposed model to discover activity and to detect behavior deviation. They perform activity discovery by binary sensor events data and then cluster the activities into original final state models using Extended Finite Automation for further activity recognition. Li et al. (2017a–c) offered sequential behavioral pattern discovery with frequent episode mining (FEM). Here FEM adapted both categorical and numerical data to mine with DBSCAN clustering algorithm. Clustering can also be applied in IoT and WSN to make sensor network even more energy efficient, optimized (Bijarbooneh et al., 2006; Xu and Helal, 2016; Heinzelman et al., 2000) and to reduce transmission distance (Choi et al., 2004; Lanzisera et al., 2014).

3.3. Association analysis or frequent pattern mining

A data object or a set of data objects or a sequences of events that appear repeatedly in a system are known as frequent patterns (Qiu et al., 2019; Che et al., 2013). Mining of these frequent patterns give a good analytical understanding of the user's activity in a felicitous environment. Recognition of pattern plays a critical role in promoting business maneuvers by mining association, correlation and other relationships among the data (Agrawal et al., 1993; Huang et al., 2004). Association Rule mining has comprehensive applications in the market basket like situations, where after analysis, one can predict the purchase pattern of the customer which can further boost the business or the user experience (Han et al., 2007; Tan et al., 2006). In the frequent pattern mining, order of events detected, matter as well. Patterns observed with specific order are called Sequential patterns and mining the same is called Sequential pattern mining (Zhao and Bhowmick, 2003). Agrawal and Srikant (1995) introduced sequential mining problem for the first time, based on the customer purchase sequence of transactions. Sequential pattern mining is preferred over Frequent pattern mining in the sensor environment as the sequence of events are observed in particular time span frequently for activity discovery and recognition (Wren and Munguia-Tapia, 2006), MavHome (Cook et al., 2003), GreaterTech smart house (Helal et al., 2005).

Technological revolution enables frequent pattern mining in environments like medical care centers or smart homes to assist in diagnosing diseases at an early stage. For example, no disease is developed in human body instantly. Rather, they spread gradually. And, the symptoms of the same can be identified by means of daily data recorded by the sensors events, the visuals or the routine check-up. Frequent pattern mining can extract useful pattern/ sequential information from the daily activities or the routine check-up data. It can thus find the deviation in user behavior and health. For e.g. if mining data for few days of an inhabitant can detect nausea, vomiting, fatigue, weakness, sleep problems, changes in urine cycle, decreased mental sharpness, muscle twitches and cramps, persistent itching, shortness of breath, high blood pressure problems, then, it will calculate the minimum Support and Confidence level to indicate that the inhabitant may be suffering from some Kidney disease (Han et al., 2007). Tsai et al. (2014) defined support and confidence as given in Eqs. (5) and (6) for Items set I = {i1, i2… ,im} and Transactions set T = {t1, t2,…, tn}.

$$\text{Support}(A => B) = P(AUB) = \frac{\chi(AUB)}{n} \tag{8}$$

$$\text{Confidence}(A => B) = P(A|B) = \frac{\chi(AUB)}{\chi(A)} \tag{9}$$

where number of transactions T that contain ψ is denoted by χ (ψ), i.e., ψ = A $\cup$ B or ψ = A. *n*denotes the total number of transactions.

Till now, we have witnessed some of the examples showing effective applications of frequent pattern mining. There are many more applications possible and served by various frequent pattern mining algorithms. A simplest well-known algorithm was proposed by Agrawal and Srikant (1994) for mining frequent patterns for Boolean association rule i.e. Apriori Algorithms. As the name itself says one should have a set of prior knowledge before applying the Apriori Algorithm. Later on, many variations of Apriori Algorithms (Park et al., 1995; Huang and Chang, 2008) were proposed to improve efficiency. Apriori achieves good performance gain but suffers when a huge number of candidate set are generated and thus increase the processing load. So, alternate methods (like FP growth, CLOSET + etc.) are proposed in (Park et al., 1995; Huang and Chang, 2008). Lee and Bang (2013) propose a track-and-trace-based anti-counterfeiting solution to discover a valid supply chain pattern for detecting counterfeit products. They first constructed a sequence tree from a trace record of EPCIS events and then applied alternative frequent pattern mining to discover valid supply chain patterns and classification to detect counterfeiting. Shukla et al. (2017) developed a system for Emotion Extractions from the text, independent of any type of dataset. They employed Class sequential rule mining which is a special case of association rule mining to extract emotions from the text. Gole and Tidke (2015) proposed hybrid method ClustBigFIMa modified version of BigFIM (Moens et al., 2013) algorithm for extracting meaningful information of association, emerging patterns, sequential pattern, correlations and other significant data mining tasks with scalability and speed. ClustBigFIM algorithm uses K-mean (Dhillon et al., 2004; Jin et al., 2006) and Apriori (Agrawal and Srikant, 1994) for generating frequent item sets and Eclat (Zaki et al., 2013) for finding potential extensions.

3.4. Other mining methods

Sometimes, exceptions faced by data mining methods are due to the unexpected useful information present in raw data known as anomalous objects or outliers (Han et al., 2007). Outlier data objects possess properties that are much different from the typical data objects. These properties may provide a good insight into some interesting inherent features. IoT applications like smart home, smart agriculture, smart traffic and parking systems, healthcare etc. benefited a lot by deviation/ outlier detection. Variety of outlier detection approaches are characterized into four categories (Tan et al., 2006) *viz.* Statistical Distribution-based outlier detection (Barnett and Lewis, 1994), Distance-based outlier detection (Knorr and Ng, 1997; Knorr and Ng, 1998), Density-based local outlier detection (Lee and Bang, 2013) and Deviation-based outlier detection (Knorr and Ng, 1997). Biswas and Misra (2015) presented a prototype of an e-health monitoring system, where they use biometric sensors and Arduino UNO board to measure and collect vital health parameters of individuals. Then they applied outlier detection mining to extract any anomalous information for health care emergencies. Yu et al. (2017) proposed outlier detection with accuracy and redundant sensor data aggregation in there cluster-based data analysis framework. They used recursive principal component analysis (R-PCA) to improve the effectiveness of IoT based systems. Zhang et al. (2018) designed a decentralized approach based on Network anomaly detection to reduce the communication overhead and achieve much faster convergence.

Furthermore, mining patterns require some other sophisticated methods too, like mining streams of data and time-series data i.e. the collection of temporal sequences. Real-time systems, communication device networks, micro-sensors devices, telemetry devices and online transactions generate streams and time series data in Exabyte volume, containing sequences of events obtained over repeated measurement of time with very fast varying update rate (Che et al., 2013). So, it is necessary for an algorithm to have one time scan, multilevel, multi-dimensional parallel real-time stream processing and analysis capability (Kesavaraj and Sukumaran, 2013). Researchers (Babcock et al., 2002; Xie et al., 2008; Roddick et al., 2002; Jensen et al., 2017) presented some good surveys on streams data mining and Time series data mining respectively. Time series data mining belongs to a Sequential Mining category which is already discussed in frequent pattern mining and sequential pattern mining section.

4. I IOT and data mining applications

4.1. J. Smart Home, Ambient Assistant Living and smart healthcare

United Nation report foretold that by the year 2050, 66% of total world population will get accommodated in an urban area (United Nation, 2014). Another study by the World Health Organization predicted that the people with age 60 and more will increase from 12% in 2015 to 22% by 2050 of the total world population (Facts and The, 2017). With increase in age, health issues increase as well, demanding more healthcare facilities and significantly increased health expenses. As the major population of the world belongs to the urban region, the best solution possible in technology geeky urban region can be a smart home and an Ambient Assistant Living (AAL) (Youngblood and Cook, 2007). Smart home and AAL, are the popular and emerging applications of IoT in the recent times. These make individual lives easier, technology friendly and even more comfortable, curable and healthier by fabricating an alive and caring environment from smart objects. Objects with embedded technologies can interact simultaneously with the other objects, individuals, internal servers and external environment (Kravchenko et al., 2017).

Human beings usually work in a cycle of 24 h. One can discover and then recognize activities and can automate them. Youngblood and Cook (2007) designed the ProPHeT decision-learning algorithm that learns strategy and controls the smart environment. They use Episode Discovery sequential pattern mining to observe activities, Active LeZi algorithm to predict upcoming action and automatically constructed hierarchical hidden Markov model that learn an action policy for the smart environment. Researchers in ref. (Samarah et al., 2017) proposed framework for wireless sensor network with spatio-temporal mining technique for activity recognition and micro-aggregation approach to enhance the privacy of user data.

References (Joergschmalenstroeer, 2010; Rashidi et al., 2011) constructed a system that recognizes and tracks user activities of daily living for a smart environment. Latter one used audio data from series of a microphone with online Diarization capability and face identification from visual data of the camera. Whereas, former introduced an unsupervised Discontinuous varied-order Sequential Miner for activity discovery and then clustered them in groups and recognized them by a boosted version of hidden Markov model. Likewise, the references (Brdiczka et al., 2007; Kasteren and Krose, 2007) used Naïve Bayes Classifier for activity recognition. This classifier yields good accuracy when data is in enormous volume with greatest probabilities to the observed values and conditional independence of the features. Li et al. (2017a–c) pointed towards the ill-suited data collection methodologies used for activity recognition and so in their research, they developed a self-constrained, scalable and energy-efficient bespoke WSN with a compact data format for episode mining to overcome this obstacle.

So far, we have discussed activity recognition and tracking in the smart environment for creating individual life easier and maximizing the comfort. This everyday activity recognition and discovery can be utilized in behavior deviation and predictions for the individuals who need medically monitored patients and healthcare emergency (Zdravevski et al., 2017; Virone et al., 2008; Yassine et al., 2017). Saives et al. (2015) proposed a binary sensor-equipped smart home to improve the autonomy of the medically monitored patient. The approach contains two sections, first uses Sequence Pattern Mining with Extended Finite Automation to model user Activities and second detects behavior deviation using residual method for any healthcare emergency. Reference (Zdravevski et al., 2017) used variety of sensor data that was segmented with sliding window, extracted time and frequency domain features with first deviation, Delta series and Fast Fourier transformation, then reduced features by Diversified Forward-Backward Feature selection and finally generate classification model with machine learning algorithms like Logistic Regression, Extended Randomize Tree and SVM with Gaussian kernel. Similarly, references (Virone et al., 2008; Yassine et al., 2017) extracted behavior deviation information for healthcare facilities in a smart environment. Where latter develop their own pattern mining software SAMCAD to mine wireless passive sensor data streams and former utilized FP-growth and k-mean mining for pattern recognition and clustering respectively and then predicted activities using Bayesian network.

All the above discussion is carried out for activity detection and deviation with respect to an individual. References (Gu et al., 2011; Alam et al., 2016) proposed systems that considered multiple individuals. Gu et al. (2011) designed a scalable and noise-resistant Emerging Pattern-based Multiuser Activity Recognizer (epMAR) with Activity Model included Emerging Pattern, Activity Correlation, Sliding window coverage. Alam et al. (2016) designed a Constraints and Correlation mining Engine (CACE) with loosely-coupled Hierarchical Dynamic Bayesian Network and data mining approach to discover key spatiotemporal constrain to prune the overall state space of coupled model. References (Li et al., 2017a–c; Rosslin and Tai-hoon, 2010) used smart home environment for helping and caring Dementia and Alzheimer patients respectively. Data mining techniques with sensors and smart equipment data can also fulfill the emerging trend towards smart energy management (Silva, 2016; Eibl et al., 2015). Silva (2016) presented unsupervised and probabilistic IPCL data fusion technique for multi-source smart home energy management. References (Angelis et al., 2013; Anvari-Moghaddam et al., 2015) proposed a Mixed-integer linear programming approach towards optimization of energy consumption as a cost reduction mechanism and to create the balance between consumption and comfort. Eibl et al. (2015) studied the impact of time granularity on edge detection methods with low-frequency Nonintrusive Appliance Load Monitoring Analyzes (NIALM) from the user privacy viewpoint.

4.2. Smart grid

Ubiquitous technological advancement due to sensor and actuator equipped embedded devices and machine dependent modern style of living has inflated the electric power demands. British Petroleum Energy Outlook 2017 (Dudley, 2017) predicted 30% increase in global energy demand by 2035 whereas United State Energy Information Administration Annual Energy Outlook 2018 (U.S. Energy Information Administration, Annual Energy Outlook, 2018) predicted for the U.S. an average of 0.4% increases in energy demand every year till 2050. This continuous inflated electric power demands, engenders huge difficulties and unbalanced situations for traditional power grid due to their characteristics. With the technological revolution, the ambient environment with Internet connectivity and communication capabilities becomes more intelligent and, thus, can measure and control power consumption interactively. This intelligence and smartness were limited to small scale applications but now can also be seen in large scale application like Smart Grid.

Smart Grid can be defined as globally existent digital grid architecture integrated with the enormous number of embedded devices that can perform a variety of digital computing task with a faster rate, precision and efficiency. The characteristics like bidirectional real-time communication, intelligently utilizing distributed power generations, distributed automatic monitoring, controlling and recovering from emergencies in real-time are integral part of it. Tuballa and Abundo (2016) presented their review from development and technological perspective, Bayindir et al. (2016) presented their review from technology and application perspective whereas Park et al. (2017) reviewed key dynamic characteristics that play a crucial role in the acceptance of smart grid technology.

With the help of information and communication technology enabled architecture, the smart grid can intelligently manage the Demand and Response of electric power. Marah and Hibaoui (2018) proposed two algorithms, first algorithm paired with branch and bound algorithm to manage domestic appliances energy consumptions as per consumer priority and regulate consumption peak to utilize it at the best. The second algorithm manages the power transmission and distribution. Rahim et al. (2018) with Hybrid Bacteria Harmony algorithm focus on demand-side management and encourage consumer with Session Time of use tariff to change consumption pattern that reduces cost and peak to average ratio. Reference (Jindal et al., 2018) introduced smart devices data-oriented consumption aware Data Analytical Demand Response management scheme for peak load reduction of residential power. Based on the demand–supply theory of economics, reference (Ferdous et al., 2017) developed an optimal dynamic pricing mechanism for trading-off between consumer utility and profit. Using this, a smart grid operator can purchase power from distributed sources and can encourage consumers with dynamic pricing. To predict power usage of the user, they applied error-back-propagation artificial neural network with feed-forward multilayer perception model.

The distributed variety of smart devices in the smart grid generates data that possess variety, velocity and volume as the inherent characteristics. This real-time big data of smart grid if utilized properly, with various big data mining methodology, can yield an efficient, reliable, sustainable and intelligent real-time monitoring and controlling. Tu et al., 2017 presented a good comprehensive big data analytic techniques and applications oriented survey on smart grid. Reference (Munshi et al., 2017) developed a cloud-supported open source Hadoop based big data framework for smart grid analytics, whereas, in (Shah et al., 2015; Yaghmaee et al., 2018) and [140] developed a summarization paradigm and Fog-Based Internet of Energy Architecture with real-time energy consumption pattern monitoring and transactive energy management respectively. This can be later used for various applications including demand-side management, Direct load control, smart pricing etc.

The massive data induced from the network of the smart device demand more precise and intelligent communication technologies. Communication infrastructure is the key factor of smart grid and this knocks the door of next-generation technologies like low power wide area network (LPWAN) enabled 5th generation mobile technology, ZigBee, WiMAX etc. or wired fiber optic, hybrid power line communication/ wireless channel (HPWC) based communication. References (Li et al., 2017a–c; Dib et al., 2018) presented a survey from both quantitative and qualitative perspective of key wireless and wired communication technologies respectively. Reference (Dib et al., 2018) suggested a Narrow Band IoT whereas in (Li et al., 2017a–c) suggested HPWC as the best solution. References (Kaur et al., 2018; Guo et al., 2017) proposed various schemes and methods to support and manage proper smart grid communication. For the sake of intelligence, the smart meter and devices data are accessed by various algorithms of smart grid infrastructure. This available data can be misused by numerous physical and cyber-attacks, therefore, it demands security and privacy preservation at all levels. References (Mahmood et al., 2017; Lyu et al., 2018; Otuoze et al., 2017; Sushmitaruj, 2013; Brunner et al., 2017) discussed and presented security and privacy requirements and solutions with highlighted problems.

4.3. Industrial IOT and smart manufacturing

The industrial revolution began in the 18th century when steam powered the industrial machines (Industry 1.0). Then came electric energy powered mass production in the 19th century (Industry 2.0) based on electronic and information technology. Industrial automation brings Industry 3.0 in the 20th century. With the advancement in Information and communication technologies, Industry 4.0 introduces machines that can connect to the network, can communicate with each other and can make decisions. Research communities like IoT, CPS, M2M communication, Additive Manufacturing and many more are persistently involved in engendering incipient standards, protocols and system architectures to destroy obstacles and empower industries in this new era of Smart Manufacturing (SM). Global economic powers like USA, Germany, Korea and China have already developed a strategic roadmap for future SM (Commission, 2010; The State Council, 2015). Properties required to achieve SM goal are Ubiquitous distributed Intelligence, uninterrupted connectivity and deeply integrated networks, Real-time knowledge creation and decision making, service-oriented standards, protocols and system architectures and knowledge oriented reconfigurable production line (Lee et al., 2014; Zhang et al., 2018; Wang et al., 2018; Ding and Jiang, 2018; Yuan et al., 2017; Thibaud et al., 2018; Alam et al., 2015a,b; Alam and El Saddik, 2017).

Authors in (Da et al., 2014; Ahuett-Garza and Kurfess, 2018; Kang et al., 2016) presented some good research reviews from various perspectives. They highlighted the key enabling technologies behind SM including identification and Tracking Technologies (RFID, smart sensors and hologram with WSN), physical wireless communication technologies (IEEE802.11x, Bluetooth, IEEE802.15.4, ZigBee and IEEE1451.5, IPv6, 4G/5G), Industrial Hardware control technologies (Distributed Control Systems, Industrial Robotics, Smart camera and imaging system, Product life cycle Management systems, Manufacturing Execution Systems and SCADA), Machine or Deep Learning, Cloudlets, Cloud manufacturing, edge computing, fog computing, Additive manufacturing, Cyber-Physical Systems, IoT and Service-oriented architectures (Alam et al., 2015a,b; Alam and El Saddik, 2017). To achieve SM, Chen et al. (2017) proposed a 4 layer hierarchical architecture for the smart factory. Four layers combine Physical Resource Layer, Network layer, Cloud Application Layer and Terminal Layer from bottom to up. Whereas, Schuh et al. (2014) deduced basic collaboration mechanism reference system on the basis of contributor enabler's analysis to empower productivity in Industry 4.0 context. The massive attempts made by machines while communication (especially in radio access networks) will lead to an intolerable delay, packet loss, and even service unavailability. To address the performance degradation due to concurrent and massive access attempts in 5G network environment, (Oh et al., 2015) proposed a joint optimal Physical Random Access Channel (PRACH) resource allocation and access control mechanism. The communication and interoperability between the heterogeneous machines plays an important role as well. Cloud based Cyber Physical Systems (C2PS) can help a lot from this aspect. The features like scalable storage, interoperability, heterogeneous computation and communication can excellently compensate by C2PS (Alam et al., 2015a,b; Alam and El Saddik, 2017). Alam et al. (2015a,b, 2017) proposed the C2PS architectures for IoT. They designed an interaction controller using a Bayesian belief network that dynamically considers current contexts and also a fuzzy logic rule base Bayes network composition for enabling reconfiguration capability.

Knowledge discovery in big data algorithms and techniques for industrial environment play the key role to transform manufacturing into smart manufacturing. Authors in reference (Lee et al., 2014) reviewed the latest trends of manufacturing for sustainable innovative services in big data environment and also show big data management with the smart predictive tool like "smart remote machinery maintenance systems with Komatsu". In ref. (Wang et al., 2018), authors submitted a comprehensive survey of commonly used deep learning algorithms. Whereas (Zhang et al., 2018) proposed big data feature learning of industrial IoT as adaptive dropout deep computational model with crowd sourcing to cloud. Smart dust to drones, smart automobiles with parking assistance and crash avoidance to smart job shop and even in hazardous situations like petrochemical plants, industrial IoT applications are producing a web across the globe (Ding and Jiang, 2018; Yuan et al., 2017) To empower these globalized industrial IoT applications, network communication, data storage and access, privacy, security and safety in the high-risk environment play a very crucial role. References (Yuan et al., 2017; Thibaud et al., 2018) presented some latest research articles addressing the above points with good insight.

4.4. Other applications

4.4.1. Smart agriculture

Steep rise in population, stagnated agriculture production, climate extremes and weather variability demand instant deployment of advanced technologies in agriculture. The quantity and quality needs of future food and nutrition supplements can only be met when both biotic and a biotic constraints are addressed with agro-technological research and development boost. With the characteristics like heterogeneous low power sensors and actuator deployment, ubiquitous network connectivity, knowledge extraction from big raw data and real-time distributed computing, IoT with cloud computing and big data mining algorithms can be the best solution for optimized and improved smart agriculture (Talavera et al., 2017; Popovic et al., 2017). The applications of Smart Agriculture are as follows: (Tzounis et al., 2017) used SmartMesh IP enabled technology to predict frost events in a peach orchard, (Severino et al., 2018; Wang, 2014) applied it for environmental risk reduction with soil moisture dynamics, Mitigation and Reduction of emission of CH_4, CO_2, NO and NO_2, water recycling, drip irrigation, hybrid machine learning based automatic plant phenotyping, smart aquaponics for urban farming and fully automated support system for non-experienced farmers.

4.4.2. Smart transportation

The racing population expansion, automobile and other routine utilities in the present and future foster significant and complex challenges that cannot be handled with current transportation management systems. Transportation today demands intelligent technologies with huge number of smart sensors and actuators, video cameras, smart cards, RFID Tags, GPS, Smartphone's, internet and social network with the big data mining techniques and intelligent control unit. The petabytes of complex data collected from these ubiquitous smart devices can be transformed into useful knowledge that have profound impact on intelligent decision making like vehicle to vehicle communication to avoid accidents and traffic congestion (Shukla et al., 2017; Saini et al., 2017; Alam et al., 2015a,b; Chowdhary et al., 2019; Tian et al., 2020a,b). Zhu et al. (2018) presented a good survey on big data analytic for intelligent transportation services (ITS). They also submitted three layers architecture of conducting big data analytics in ITS *viz.* data collection, analytics and application layer. The smart transportation applications are the integral part of smart city application where numerous other field applications devices can also participate to build smarter transportation. And so there are several ways to acquire verity of data (Shukla et al., 2017). A survey has been conducted by (Shukla et al., 2017) on the ways of data collection by the mobile smart vehicle. Vehicles are accessing and generating data, either of entertainment type or of informative type (Saini et al., 2017). Extracting the useful information from it requires a real time solution. Saini et al. (2017) presented "InCloud" framework, a cloud base middleware solution for vehicular infotainment application development. Chowdhary et al. (2019, 2020a,b) Presented a IoT based battery health monitoring and alert system for vehicle to avoid malfunctioning condition. To overcome huge data from an embedded lightweight Clint application installed with internet connected vehicle, they incorporated data filtering and fusion functionalities. Not only the infotainment data but vehicle can also exchange several services data (Like safety, comfort and efficiency) for social good. The Vehicular Ad-hoc Network (VANETs) plays a very important role here. Alam et al. (2015a,b) presented architecture that support vehicle to vehicle, vehicle to infrastructure and vehicle to internet communication based on cloud oriented Cyber Physical Social Internet of Vehicle (SIoV). It utilizes the social relationships among the physical components (instead of device owners) for social services, although the communication between physical devices without knowing to the device owner will also create several threats for security and privacy. Passing a single wrong message may lead to a threat for lives. Tian et al. (2020a,b) proposed a trust management framework "Vcash" for Internet of Vehicles. Their mechanism identifies malicious vehicle and also encourage vehicle to provide qualified sensed traffic events. There were several applications of Smart Transportation system where data mining play a vital role. These include: utilizing multivariate logistic regression algorithm for fatality risk detection for driver (Bédard et al., 2002), Bayesian inference and Random forest for real-time crash detection, Two Stage Divide & Conquer (TSDC) algorithm/ Branch and Bound (BB) algorithm/ Dijkstra's algorithm to find out the shortest path (Katre et al., 2017). Fog-FISVER for real-time crime detection on public bus services (Neto et al., 2018), Dynamic Bayesian Network for Fatigue Modeling and for categorizing human fatigue expressions and driver distraction detection, cascade classifiers for face detection (Kaplan et al., 2015), multivariate statistical model for weather prediction, hierarchical tree-based regression and binomial regression model for accident detection and emergency services (Zhu et al., 2018) and many more. There are many advanced applications like Traffic control and future predictions, Intelligent transportation planning and execution, logistics, automated driving, automatic and accurate identification of security threats with smart transportation safety, smart parking and vehicle communication are the future research challenges where one can focus (Shukla et al., 2017; Saini et al., 2017; Katre et al., 2017; Brunner et al., 2017; Bédard et al., 2002; Alam et al., 2015a,b).

Table 1
Smart Home, AAL and Smart Healthcare Application Objective, and Applied Mining Algorithms.

Application	Objectives	Raw Data Sources	Data mining algorithms used for knowledge conversion			
			Classification	Clustering	Frequent and Sequential Pattern Mining	Other Mining Algorithms
Smart Home, Ambient Assistant Living and Smart Healthcare	• Activity Discovery, Identification, recognition and Prediction (for single user) (Youngblood and Cook, 2007; Samarah et al., 2017; Joergschmalenstroeer, 2010; Rashidi et al., 2011; Saives et al., 2015) and for multiple users (Gu et al., 2011; Alam et al., 2016) • improve quality and comfort of life, energy conservation, privacy, safety and security (Youngblood and Cook, 2007; Samarah et al., 2017; Joergschmalenstroeer, 2010; Rashidi et al., 2011; Saives et al., 2015; Silva, 2016; Angelis et al., 2013; Anvari-Moghaddam et al., 2015; Eibl et al., 2015) • Spasticity/ resuscitation/ health care emergency Detection (Biswas and Misra, 2015; Li et al., 2017a–c; Zdravevski et al., 2017; Virone et al., 2008; Yassine et al., 2017; Li et al., 2017a–c; Rosslin and Tai-hoon, 2010) • Emotions Extractions or Behavior pattern deviation detection (Gole and Tidke, 2015; Li et al., 2017a–c; Zdravevski et al., 2017; Virone et al., 2008; Yassine et al., 2017) • Support system for Dementia and Alzheimer patients (Li et al., 2017a–c; Rosslin and Tai-hoon, 2010)	• Text data • Series of Microphones and video cameras • Wireless sensors • Wearable sensors • Binary sensors • Bespoke WSN • Smart home appliance data • Smart energy meter data • Smart phone	• KNN • hierarchical HMM, Boosted HMM • Naïve Bayes • Logistic Regression • Extended Randomize Tree • SVM with Gaussian kernel • Loosely-coupled Hierarchical Dynamic Bayesian network	• k-mean • k-anonymity • DBSCAN • Extended Finite Automation • Micro-aggregation	• FP-growth • Episode Discovery sequential pattern mining • Frequent episode mining • unsupervised Discontinuous varied-order Sequential Miner • Extended Finite Automation	• Residual method • Unsupervised and probabilistic IPCL data fusion technique • eMAR (Gu et al., 2011) • CACE (Alam et al., 2016)

5. Summary and open research issues in IoT applications

As per the literature reviewed, Tables 1–6 summarize various objectives of IoT infrastructure, data sources and the utilized data mining algorithms to extract knowledge. This article significantly strengthens researchers and developers aiming for fabricating an interactive environment for the numerous prospective IoT applications. IoT and cloud-assisted data mining technologies equipped with sensors and actuators can drastically transform the passive environment into an intelligent and active environment. In the following text, we summarize and point wise discuss open research issues from diverse IoT applications' outlook.

Smart Home, AAL and Smart Healthcare: Research on the intelligent buildings with smart home, AAL and healthcare applications, majorly focuses on activity discovery, behavior change recognition, activity prediction, optimization and emergency detection. The intelligent building IoT architecture, extracts information from plethora of raw data and prepares analytics with the help of data mining algorithms specific to the given IoT applications task. It gives feedback to the system to provide exceptional comfort and quality of living, energy management, and establishes a support system for the patients.

Open Research Issues:

- Smart devices (as specified in the raw data sources column of Table 1) in IoT environment learn daily activities from real-time information and then manage system according to the personal pattern. Therefore robust and flawless device integration and management capabilities are required.
- The diverse components should exchange information and support each other intelligently, irrespective of whether they were designed to work together or not. For e.g., healthcare applications can utilize and control HVAC home appliances but energy management units should not turn off healthcare devices for energy saving purposes. Consecutively, these applications require middleware solution for interoperability.
- Devices connected to the smart house/ building network use various protocols like NFC, Bluetooth, DASH7, ZigBee, Wi-Fi, 4G, 5G and many more. The key factors that should be considered here is the network latency and the available bandwidth as it varies with the protocols. It can highly affect the precision in time-sensitive restricted environment.
- Heterogeneous big IoT data storing, processing and knowledge conversion is another key feature that needs to be addressed. A context aware parallel processing architecture should be imposed with multiple data mining algorithms, to apply a particular or an ordered set of algorithms for serving dynamic application environment.
- The service needs change person to person and even according to the work environment. So, the research in intelligent building should focus on service oriented self-adoption mechanism that can update itself when required.
- As the number of devices communicates, they generate issues like privacy of personal data and physical location tracking. Unsafe device behavior, unauthorized access and service modification can create glitches to the security in any intelligent building environment. The gaps between the security and privacy can be minimized with suitable machine learning researches.

Smart Grid and Smart manufacturing: The development of smart grid and smart industries is a comprehensive process of complexity and competence among human to machine and machine to machine. The research under Smart grid concentrates on the management of distributed power generation (like renewable and non-renewable resources), Demand-side, peak hour power & home appliances utilization, automated revenue collection, smart pricing, communication over power line, fault detection and recovery. Whereas the Smart manufacturing converges on automatic identification, monitoring & controlling of machines, logistics management, Emergency/ Disaster /Safety

Table 2
Smart Grid Objective and Applied Mining Algorithms.

Application	Objectives	Raw Data Sources	Data mining algorithms used for knowledge conversion			
			Classification	Clustering	Frequent and Sequential Pattern Mining	Other Mining Algorithms
Smart Grid	• Automated meter reading (Marah and Hibaoui, 2018; Munshi et al., 2017) • Peak hour Power utilization, quality and Demand side management with load balancing (Bayindir et al., 2016) (Marah and Hibaoui, 2018; Rahim et al., 2018; Jindal et al., 2018; Munshi et al., 2017; Shah et al., 2015; Yaghmaee et al., 2018; Li et al., 2017a–c) • Smart domestic appliance management (Marah and Hibaoui, 2018; Rahim et al., 2018; Jindal et al., 2018) • Always on access with real time pricing and bidirectional real-time communication (Marah and Hibaoui, 2018; Rahim et al., 2018; Jindal et al., 2018; Ferdous et al., 2017; Tu et al., 2017; Munshi et al., 2017; Shah et al., 2015; Yaghmaee et al., 2018; Li et al., 2017a–c; Dib et al., 2018; Kaur et al., 2018; Mahmood et al., 2017) • Smartly utilizing distributed power generations like renewable energy (Ferdous et al., 2017; Tu et al., 2017; Munshi et al., 2017; Shah et al., 2015; Yaghmaee et al., 2018; Li et al., 2017a–c; Dib et al., 2018; Kaur et al., 2018) • Smart fault detection and real-time recovery (Tu et al., 2017), Security and Privacy preservation (Mahmood et al., 2017; Lyu et al., 2018; Otuoze et al., 2017)	• Smart meters • RFID tags and readers • WiFi enable WSN equipped utilities like smart appliances, smart power generator and transmission devices. • Renewable energy sources data like solar panel, wind turbine etc.	• Error-back-propagation artificial neural network with feed forward multilayer perception model • Scatter-plot-based event classification • Decision tree and random forest	• Hieratical, portioning and density based approach • K-mean • fuzzy c-mean • eXtended classifier system for clustering • Tensor based data management algorithm	• CC-DADR and UC-DADR(Jindal et al., 2018) • All frequent and sequential mining algorithms of smart home can be utilized for smart energy management	• Branch and bound algorithm • Distributed Max-Flow algorithm • Metaheuristic hybrid bacteria harmony algorithm • Summarization algorithm with stream data processing (Shah et al., 2015)

management, Automatic Fault diagnosis, raise in productivity, serviceable life prediction of product and cloud-based manufacturing automation.

Open Research Issues:

- Advancement in sensing and actuation has produced ubiquitous computing environment from an LED bulb to a full-fledged automated machine. Therefore, a vigorous research is required on a broader scale for reliable, energy efficient and flexible communication protocol for use in highly heterogeneous and resource constraint smart grid.
- The existing infrastructure for both in power generation and manufacturing is inadequate, especially in developing countries. The issues like clean and distributed energy generation and management, framework for utilizing the potential of participatory sensing, bidirectional power flow, self-reconfigurable machines, and support for generic M2M communication protocols need to be focused upon.
- The existing technologies ranging from digital to advanced sustainable manufacturing in a connected environment should be integrated with cloud based machine learning capabilities that can adopt new algorithms (especially Metaheuristic and greedy algorithms for unstructured data) with existing realities. This will benefit the end user and several SCADA/Discrete control systems with generated analytics
- Resources in both the applications field suffer several restrictions. Physical resources should be integrated with the capabilities like efficient data acquisition, analytics visibility to end user and extensible controller access to the core industrial network.
- With the growing trends, the equipment and systems are getting obsolete. Strategic research is required for system lifecycle management. This includes designing and manufacturing goals, maintenance goals, and an assisting model for interoperability between the old and new generation devices having huge variation from design to standards and protocols.
- Due to non-uniform power generation in distributed grid, it demands huge storage capacity. Most of the batteries have very short life span of a few years, are heavy and of large size. Research over smarter and lighter batteries while covering all above aspects is one of the highlighted fields in smart grid.
- Information security and privacy are extremely vital for power and manufacturing utilities, especially for billing purposes, grid control and manufacturing/production data. To avoid cyber-attacks, to prevent unauthorized access to connected control systems and to

Table 3
Industrial IOT and Smart Manufacturing Objective and Applied Mining Algorithms.

Application	Objectives	Raw Data Sources	Data mining algorithms used for knowledge conversion			
			Classification	Clustering	Frequent and Sequential Pattern Mining	Other Mining Algorithms
Industrial IOT and Smart	Manufacturing	• Detecting counterfeit products (Lee and Bang, 2013) • Real-time Automatic Identification, monitoring and controlling manufacturing objects (Da et al., 2014; Wang et al., 2018; Ding and Jiang, 2018) • Food supply chain: quality, quantity, efficiency and food safety management (Da et al., 2014; Thibaud et al., 2018) • Disaster and safety management in critical manufacturing conditions (Ahuett-Garza and Kurfess, 2018), smart maintenance (Chen et al., 2017; Lee et al., 2014; Wang et al., 2018), • Intelligent oil refinery and petrochemical industry (Shu et al., 2018; Yuan et al., 2017) • Fault diagnosis, raise productivity and useful life prediction of product (Ahuett-Garza and Kurfess, 2018; Schuh et al., 2014) • Cloud based knowledge driven smart manufacturing automation or Lean automation (Kang et al., 2016; Chen et al., 2017; Wang et al., 2018; Zhang et al., 2018)	• RFID, PLC • smart sensors • Bar-code, hologram with WSN • Distributed Control Systems • Smart Industrial Robotics • Smart camera and imagine system • Manufacturing Execution Systems	• Regular expression technique • Artificial neural network • Adaptive neuro-fuzzy interface system • Hierarchical/convolutional deep neural network • Apriori algorithm • Restricted Boltzmann machine its variant • Recurrent neural network its variant • SLME (Zhang et al., 2018), SVM • Principle component analysis	• k-mean • C 4.5 • C 5.0 • Shape mining • Self organized Map and Gaussian Mixture model • Expectation	Maximization • ID-AVL Tree and RF Tree

(continued on next page)

Table 3 (*continued*)

Application	Objectives	Raw Data Sources	Data mining algorithms used for knowledge conversion			
			Classification	Clustering	Frequent and Sequential Pattern Mining	Other Mining Algorithms
		• Energy efficient resource allocation and advanced communication framework (Yuan et al., 2017; Thibaud et al., 2018; Alam et al., 2015a,b; Alam and El Saddik, 2017; Tzounis et al., 2017)				
• Alternative frequent pattern mining • Auto encoder and its variant • Deep belief network	• Collaboration Mechanism • PHM Algorithm (Lee et al., 2014) • LPT-DP-K algorithm • Laplace based noise enhancement algorithm • Weighted selection based on exponential Mechanism (Tzounis et al., 2017)					

avoid misuse of enterprise data in high level of connectivity, an efficient security mechanism should be developed. Standardization efforts regarding the security and privacy should be made.

Smart agriculture and Smart transportation: IoT is gradually transforming the human operated agriculture and transportation environment to self-organized automated environment. The research in the smart agriculture domain include the climate variation monitoring, forecasting and optimization, environment predictions, alerting and risk reduction system, keenly intellective farming, rainwater harvesting, automatic plant phenotyping and aquaponics urban farming. Smart transportation research objectives include traffic forecasting, intelligent transportation service planning, accident avoidance and emergency support system, train and air traffic management, driver behavior analysis, smart parking and smart automobiles. Both of these application areas are dealing with large geographical region. While the former has majorly stable embedded nodes spread over the field and the latter has majorly mobile nodes with ever changing density.

Open Research Issues:

- Devices in both the application environments handle a harsh environment with limited power supply. A smart and efficient power utilization system architecture must be applied that can program short time slot activation of a devices without missing any important information.
- Huge number of sensors and actuators are involved in the network. Heterogeneity in devices, standards, protocols, generated data, memory, processing and power constraints are handled by the network dynamically. Lot of good research work (like LEACH and its variants) in this field with WSN has already been published.
- Data gathering is also one of the biggest challenges in a dynamic environment. There are various types of data like random logistic and traffic data, vehicle to vehicle communication data, road side unit's data, environmental monitoring data, linguistic data, GPS data and many more. The devices occupy diverse methodologies to gather this heterogeneous data. Applying mining technologies to extract useful analytics (like driver behavior recognition, linguistic identification, automated car support etc.) and to prepare them in several formats to provide services to highly divergent devices is an emerging research area.
- The plethora of data generated in the highly resource constraint agriculture and transportation application environment can be supported well by Cloud based infrastructure. Other services include high quality analytics, hardware support, interoperability, storage and computational resources to process the data at the edge of the network.

In summary, an IoT system architecture provides services to several applications. It is loaded with the massive data generated from the intellective sensors and the embedded devices. The big data transforms into valuable knowledge with data mining techniques, which later can be utilized for even more intelligent resources and service management. Advanced classification and clustering techniques are the frequently used data mining algorithms. These include SVM, HMM, Bayesian network, logistic regression, deep neural network, k-mean, DBSCAN, Extended Finite automation, fuzzy c-meant. Frequent and Sequential Pattern mining algorithms like FP-growth, Episode Discovery, Varied-order Sequential Miner, Deep belief network and variants play a significant role when the sequential real-time events or patterns are discovered. Future IoT infrastructure requires a lot of in-depth research and development integrated with machine learning capabilities.

As depicted in Fig. 3, in high-performance IoT environment, applications will be benefited with the integrated multiple data mining technologies. In most of the applications, classification and clustering algorithms are used interchangeably or coherently for the complex data set. Some applications perform clustering then classification, some do classification before clustering. Whereas some applications (Youngblood and Cook, 2007; Alam et al., 2016; Jindal et al., 2018; Zhu et al., 2018) begin with frequent/sequential pattern mining or outlier detection and then the extracted information is supplied to the clustering and then to the classification algorithm. By repeatedly employing several data mining algorithms in sequence, applications can produce and utilize distinct classifiers, clusters or even can find unpredicted important frequent/sequential

Table 4
Smart Agriculture Objective and Applied Mining Algorithms.

Application	Objectives	Raw Data Sources	Data mining algorithms used for knowledge conversion		
			Classification	Clustering	Other Mining Algorithms
Smart	Agriculture	• Rainfall Forecasting (Wu, 2009) • Climate monitoring and optimization (Tzounis et al., 2017), smart greenhouse gas emission control (Talavera et al., 2017; Popovic et al., 2017) • Environment predictions, Alerting and risk reduction system (Tzounis et al., 2017; Popovic et al., 2017) • Agricultural field, fertilizers and pesticide monitoring and control (Tzounis et al., 2017; Popovic et al., 2017) • Cloud based Intelligent automated irrigation system (Tzounis et al., 2017; Severino et al., 2018; Wang, 2014) • Food supply chain management (Da et al., 2014; Talavera et al., 2017; Kyaw et al., 2017) • Predict frost events in a preach orchad (Popovic et al., 2017) • Rain water harvesting, drip irrigation (Severino et al., 2018) • Automatic plant phenotyping (Yahata et al., 2017) • Smart aquaponics urban farming (Kyaw et al., 2017)	• GIS system • RFID, NFS • Optical sensors • Low power wireless sensor network • Smart agriculture sensors and devices • Video camera and processing tool • Solar panel • Web data • Human as sensor	• Support Vector Machine • Scale vector machine • Random forest • Extremely randomized tree • Decision tree • Artificial neural network • Logistic regression • Bayesian belief network • Naïve Bayes • Convolutional neural network	• K-mean • Farthest first clustering algorithm • Spatial Fuzzy clustering algorithm • Simple linear iterative clustering
• Statistical analysis • GIS geospatial analysis • Image processing • NVDI vegetation Indices • Branch and bound enumeration					

Table 5
Smart Transportation Objective and Applied Mining Algorithms.

Application	Objectives	Raw Data Sources	Data mining algorithms used for knowledge conversion		
			Classification	Clustering	Other Mining Algorithms
Smart	transportation	• Traffic Sign Detection and Recognition (Abedin et al., 2017) • Traffic predictions and forecasting (Zhu et al., 2018; Shukla et al., 2017; Alam et al., 2015a, b) • Identification, monitoring and controlling logistics (Da et al., 2014; Shukla et al., 2017) • Transportation service planning (Zhu et al., 2018; Alam et al., 2015a, b) • Smart automobiles (Shukla et al., 2017; Saini et al., 2017) • Vehicle to vehicle communication (Alam et al., 2015a,b; Alam and El Saddik, 2017) • Road traffic accident analysis and emergency supporting system (Zhu et al., 2018; Shukla et al., 2017; Saini et al., 2017) • Smart train management (Zhu et al., 2018) • Driver behavior analysis (Saini et al., 2017) • Automated driving (Tian et al., 2020a) • Automatic and accurate identification of security threats with smart transporta-	• RFID, GPS • Smart card • Smart automobiles • Road side smart sensors and actuators • Camera, GIS • Mobile phones • Social media • Web data • Smart wearable sensors • Vehicle sensor devices	• Linear regression • Hierarchicaltree-based regression • Decision trees • Convolutional neural networks support vector machines • Restricted Boltzmann Machine • Recurrent or deep Neural Network • Haar-like, HOSVD[188] • Naïve Bayes • Bayesian belief network (Alam et al., 2015a,b; Alam and El Saddik, 2017)	• K-mean • Markov chain model • Gaussian process regression algorithm • Divide and conquer algorithmALT algorithm

Table 5 (*continued*)

Application	Objectives	Raw Data Sources	Data mining algorithms used for knowledge conversion		
			Classification	Clustering	Other Mining Algorithms
		tion safety (Kaplan et al., 2015) • Smart Parking(Katre et al., 2017)			
• Reinforcement learning • Discrete traffic state encoding • Q-Learning • Stacked	autoencoders • Polynomial adaline algorithm • Heuristic online learningalgorithm Gradient-descent based online learning algorithm • RL-based online learning algorithm • Linear Discriminate Analysis (LDA) • FFT, DWT • Dijkstra's, A*, Floyd–Warshall, Johnson Algorithm • Shortest path finding algorithm				

Table 6
Other Applications Objective and Applied Mining Algorithms.

Application	Objectives	Raw Data Sources	Data mining algorithms used for knowledge conversion		
			Classification	Clustering	Other Mining Algorithms
Miscellaneous	applications	• Data fusion and recovery for resource limited smart devices	(Bijarbooneh et al., 2006) • IoT devices scalability management (Xu and Helal, 2016) • Energy efficient and optimized sensor network (Bijarbooneh et al., 2006; Xu and Helal, 2016; Heinzelman et al., 2000; Yu et al., 2017) • Shorter transmission distance (Choi et al., 2004; Lanzisera et al., 2014)	• RFID Tags, Smart Card, GPS and GIS • Low power wireless sensor network • Audio and visual devices • Other smart devices etc.	• Event driven service oriented Architecture • Lightweight quarter-sphere SVM
• Heuristic based greedy algorithm • LEACH	(Heinzelman et al., 2000) • Two phase clustering (TPC) • Scheme in multihop WSN • Conventional Data Aggregation algorithms • R-PCA based Outlier Detection Algorithm(Yu et al., 2017) • Decentralized Riemannian clusteralgorithm	• Multiphase adaptive algorithm with belief	propagation • R-PCA with SPE score based Outlier Diagnosis Algorithm (Yu et al., 2017)		

information from huge heterogeneous data of several devices. Therefore, the cloud-assisted IoT architecture integrated with deep learning on data mining technologies is the best solution for today's smart service-oriented environment.

6. Conclusion

The paper presented a systematic and detailed review of data mining algorithms like classification, clustering, and frequent/sequential pattern mining from IoT applications' perspective. The article summarized them in tabular form with the open research issues. We descriptively analyzed the applications like Smart Home, Ambient Assistant Living, Smart Healthcare, Smart Grid, Industrial IoT and Smart Manufacturing, Smart Agriculture and Smart transportation on the basis of data mining technologies employed for data to knowledge conversion. This conversion increases complexity and intellect into today's huge data producing IoT environment.

We also presented an overview of cloud assisted system architecture in Section 2 and Data mining process in Section 5, for IoT which shows that pre-processing and knowledge discovery plays the most important role among all the six layers. For systems that contain various dissimilar smart devices, and that produce heterogeneous data, pre-processing is of great importance. Knowledge discovery boosts the system performance by more appropriate and advanced service suggestions.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abedin, Z., Dhar, P, Hossenand, M. K., Deb, K., 2017. Traffic Sign Detection and Recognition Using Fuzzy Segmentation Approach and Artificial Neural Network Classifier Respectively. International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 518-523, doi: 10.1109/ECACE.2017.7912960.

Agrawal R., Srikant, R., 1994. Fast algorithms for mining association rule. in Proc. 1994 International conference on Very Large Data Base (VLDB'94), pp. 487-499.

Agrawal, R., Imielinski, T., Swami A., 1993. Mining association rules between sets of items in large databases. in Proc. ACM SIGMOD International Conference on Management of Data. (22)2, 207–216.

Agrawal, R., Srikant, R., 1995. Mining Sequential Patterns. in Proc. International conference on Data Engineering (ICDE'95), Taipei, Taiwan, Mar 1995, pp. 3–14.

Ahuett-Garza, H., Kurfess, T., 2018. A brief discussion on the trends of habilitating technologies for Industry 4.0 and Smart Manufacturing. Manufacturing Letters (2018), (15), 60-63. doi: https://doi.org/10.1016/j.mfglet.2018.02.011.

Alam, K. M., El Saddik, A., 2017. C2PS: A Digital Twin Architecture Reference Model for the Cloud-Based Cyber-Physical Systems. in IEEE Access. (5), 2050-2062. doi: 10.1109/ACCESS.2017.2657006.

Alam, K. M., Saini, M., Saddik, A. E., 2015. Toward Social Internet of Vehicles: Concept, Architecture, and Applications. in IEEE Access. (3), 343-357. doi: 10.1109/ACCESS.2015.2416657.

Alam, K. M., Sopena, A., Saddik, A. E., 2015. Design and Development of a Cloud Based Cyber-Physical Architecture for the Internet-of-Things. IEEE International Symposium on Multimedia (ISM), Miami, pp. 459-464, doi: 10.1109/ISM.2015.96.

Alam, M. A., et al., 2016. CACE: Exploiting Behavioral Interaction for Improved Activity Recognition in Multi-Inhabitant Smart Home. in proc. IEEE 36th International Conference on Distributed Computing Systems, Nara, Japan, pp. 539-548, doi: 10.1109/ICDCS.2016.61.

Alsabti, K., Ranka, S., Singh, V., 1998. CLOUDS: A Decision Tree Classifier for Large Datasets. in Proc. of the 4th Intl. Conf. on Knowledge Discovery and Data Mining, New York, pp. 2-8.

Angelis, F.D., Boaro, M., Fuselli, D., Squartini, S., Piazza, F., Wei, Q., 2013. Optimal home energy management under dynamic electrical and thermal constraints Industrial Informatics. IEEE Trans. 9, 1518–1527.

Ankerst, M., Breuning, M., Kreigel, H. P., Sander, J., 1999. OPTICS: Ordering points to identify the clustering structure. in proc. ACM-SIGMOD International Conference Management of Data (SIGMOD'99), pp. 49-60.

Anvari-Moghaddam, A., Monsef, H., Rahimi-Kian, A., 2015. Optimal smart home energy management considering energy saving and a comfortable lifestyle. Smart Grid. IEEE Trans. 6, 324–332.

Babcock, B., et al., 2002. Models and issues in data stream systems. in proc. 2002 ACM Symposium on Principles of Database Systems (PODS'02), pp. 1-16.

Bandyopadhyay, D., Sen, J., 2011. Internet of things: applications and challenges in technology and standard. Wireless Pers. Commun. 58 (1), 49–69.

Barnett, V., Lewis, T., 1994. Outliers in Statistical Data. John Wiley & Sons.

Bayindir, R. et al., 2016. Smart grid technologies and applications. Renew. Sustain. Energy Rev. 66, 499–516.

Bédard, M., Guyatt, G. H., Stones, M. J., Hirdes, J. P., 2002. The independent contribution of driver, crash, and vehicle characteristics to driver fatalities. Accident Anal. Prevention. (34)6, 717–727.

Berry, M. W., Browne, M., 2005. Understanding Search Engines: Mathematical Modeling and Text Retrieval. (17), SIAM.

Bijarbooneh, F. H., Du, W., H-Ngai, E. C., et.al, 2006. Cloud-Assisted Data Fusion and Sensor Selection for Internet of Things. IEEE Internet of Things Journal (3)3, 257-268.

Biswas, S., Misra, S., 2015. Designing of a Prototype of e-Health Monitoring System. IEEE International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Kolkata, India, Nov. 2015.

Brdiczka, O., Reignier, P., Crowley, J.L., 2007. Detecting Individual Activities from Video in a Smart Home. in Proc. 11th Int'l Conf. Knowledge-Based and Intelligent Information and Eng. Systems (KES), pp. 363–370.

Brunner, S. et al., 2017. Ontologies Used in Robotics: A Survey with an Outlook for Automated Driving. IEEE International Conference on Vehicular Electronics and Safety (ICVES), Vienna, Austria. pp. 81-84, doi: 10.1109/ICVES.2017.7991905.

Che D., Safran M., Peng Z., 2013. From Big Data to Big Data Mining: Challenges, Issues, and Opportunities. In: Hong B., Meng X., Chen L., Winiwarter W., Song W. (eds) Database Systems for Advanced Applications, DASFAA. Lecture Notes in Computer Science, (7827), pp. 1-15 2013. 10.1007/978-3-642-40270-8_1.

Chen, B. et al., 2017. Smart factory of industry 4.0: key technologies, application case, and challenges. IEEE Access 6, 6505–6519.

Chen, Q. et al., 2019. A survey on an emerging area: deep learning for smart city data. IEEE Trans. Emerg. Top. Comput. Intell. 3 (5), 392–410. https://doi.org/10.1109/TETCI.2019.2907718.

Choi, W., Shah, P., Das, S. K., 2004. A framework for energy-saving data gathering using two-phase clustering in wireless sensor networks. in Proc. International Conference on Mobile and Ubiquitous Systems, pp. 203–212.

Chowdhary, R. R., Chattopadhyay, M. K., Kamal, R., 2020. IoT based State of Charge and temperature monitoring system for mobile robots. In: Saini H., Singh R., Tariq Beg M., Sahambi J. (eds) Innovations in Electronics and Communication Engineering. Lecture Notes in Networks and Systems. (107), pp.401-413.

Chowdhary, R. R., Chattopadhyay, M. K., Kamal, R., 2020. Orchestrator Controlled Navigation of Mobile Robots in a Static Environment. In: Saini H., Singh R., Tariq Beg M., Sahambi J. (eds) Innovations in Electronics and Communication Engineering. Lecture Notes in Networks and Systems. (107), pp.193-206.

Chowdhary, R.R., Sunhare, P., Chattopadhyay, M.K., Kamal, R., 2019. IoT model based battery temperature and health monitoring system using electric vehicle like mobile robot. J. Adv. Robot. 6 (3), 1–13.

Commission, E., 2010. Europe 2020: A Strategy for smart, sustainable and inclusive growth. Working paper {COM (2010) 2020}.

Cook, D., Youngblood, M., Heierman, I., . Gopalratnam, E.O.K, Rao, S., Litvin, A., Khawaja, F., 2003. Mavhome: An Agent-Based Smart Home. in Proc. First IEEE Int'l Conf. Pervasive Computing and Communication, pp. 521–524.

Da, Li, He, W., Li, S., 2014. Internet of things in industries: a survey. IEEE Trans. Indus. Informat. 10 (4), 2233–2243. https://doi.org/10.1109/TII.2014.2300753.

Dai, H., Zheng, Z., Zhang, Y., 2019. Blockchain for internet of things: a survey. IEEE Int. Things J. 6 (5), 8076–8094. https://doi.org/10.1109/JIOT.2019.2920987.

Dhillon, I.S., Guan, Y., Kernel, K.B., 2004. k-means: spectral clustering and normalized cuts. KDD 2004, 551–556.

Ding, K., Jiang, P., 2018. RFID-based production data analysis in an IoT-enabled smart job-shop. IEEE/CAA J. Automat. Sin. 5 (1), 128–138.

Dudley, B., "BP Energy Outlook 2017," British Petroleum global, Available: https://www.bp.com/content/dam/bp/pdf/energy-economics/energy-outlook-2017/bp-energy-outlook-2017.pdf.

Eibl, G. et al., 2015. Influence of data granularity on smart meter privacy. IEEE Trans. Smart Grid 6 (2), 930–939. https://doi.org/10.1109/TSG.2014.2376613.

Ester, Kreigel, H. P., Sander, J. Xu, X., 1996. A density based algorithm for discovering clusters in large spatial databases. in proc. International conference on Knowledge Discovery and Data Mining (KDD'96), Portland, OR, pp. 226-231.

FACTS ON AGEING AND THE LIFE COURSE. 2017. Available: http://www.who.int/features/factfiles/ageing/ageing_facts/en/.

Feng Chen et al., 2015. Data Mining for the Internet of Things: Literature Review and Challenges. Hindawi Publishing Corporation International Journal of Distributed Sensor Networks.

Ferdous, J., et al., 2017. Optimal Dynamic Pricing for Trading-Off User Utility and Operator Profit in Smart Grid IEEE Transactions On Systems, Man, And Cybernetics: Systems. (50)2, 455-467. doi: 10.1109/TSMC.2017.2764442.

Fleury, A., Vacher, M., Noury, N., 2010. SVM-based multimodal classification of activities of daily living in health smart homes: sensors, algorithms, and first experimental results. IEEE Trans. Inf. Technol. Biomed. 14 (2), 274–283.

Ganz, F., Puschmann, D., Barnaghi, P., Carrez, F., 2015. A practical evaluation of information processing and abstraction techniques for the internet of things. IEEE Internet Things J. 2 (4), 340–354.

Geepalla, E., Bordbar, B., Du, X., 2013. Spatio-temporal Role Based Access Control for Physical Access Control Systems. Fourth International Conference on Emerging Security Technologies, Cambridge 2013, 39-42. doi: 10.1109/EST.2013.13.

Gole, S., Tidke, B., 2015. Frequent Itemset Mining for Big Data in social media using ClustBigFIM algorithm. International Conference on Pervasive Computing (ICPC), Jan. 2015, pp. 1-6, doi: 10.1109/PERVASIVE.2015.7087122.

Google, 2017. What is big data. [online]. Available: https://cloud.google.com/what-is-big-data/.

Gu, T. et al., 2011. Recognizing multiuser activities using wireless body sensor networks. IEEE Trans. Mobile Comput. 10 (11), 1618–1631. https://doi.org/10.1109/TMC.2011.43.

Guha, S., Rastogi, R., Shim, K., 1998. CURE: an efficient clustering algorithm for large databases. ACM SIGMOD Record. 27 (2), 73–84.

Guha, S., Rastogi, R., Shim, K., 1999. ROCK: a robust clustering algorithm for categorical attributes. in Proceedings of the 15th International Conference on Data Engineering (ICD '99), pp. 512–521.

H. Guo, H., Liu, J., Jhao, L., 2017. Big Data Acquisition under Failures in fiwi Enhanced Smart Grid. IEEE Transactions on Emerging Topics in Computing, Early Access. (7)3, 420-432. doi: 10.1109/TETC.2017.2675911.

Han, J., Cheng, H., Xin, D., Yan, X., 2007. Frequent pattern mining: current status and future directions. Data Mining Knowl. Discov. 15 (1), 55–86.

Heinzelman, W. R., Chandrakasan, A., Balakrishnan, H., 2000. Energy-efficient communication protocol for wireless micro-sensor networks. in Proc. Hawaii International Conference on System Sciences. (2), pp. 10. doi: 10.1109/HICSS.2000.926982.

Helal, S. et al., 2005. The gator tech smart house: a programmable pervasive space. Computer. 38 (3), 50–60.

Hinneburg, A., Keim, D. A., 1998. An efficient approach to clustering in large multimedia databases with noise. in proc. International Conference on Knowledge Discovery and Data Mining (KDD'98), pp. 58-65.

Huang, K.Y., Chang, C.H., 2008. Efficient mining of frequent episodes fromcomplex sequences. Inform. Syst. 33 (1), 96–114.

Huang, K., Chang, C., Lin, K., 2004.. Prowl: an efficient frequent continuity mining algorithm on event sequences. in Data Warehousing and Knowledge Discovery, Lecture Notes in Computer Science, (3181) ,pp. 351–360.

Jensen, S.K. et al., 2017. Time series management systems: a survey. IEEE Trans. Knowl. Data Eng. 29 (11), 2581–2600. https://doi.org/10.1109/TKDE.2017.2740932.

Jin, R., Goswami, A., Agrawal, G., 2006. Fast and exact out-of-core and distributed k-means clustering. KnowlInfSyst 10 (1), 17–40.

Jindal, A. et al., 2018. Consumption-aware data analytical demand response scheme for peak load reduction in smart grid. IEEE Trans. Ind. Electron. 65 (11), 8993–9004. https://doi.org/10.1109/TIE.2018.2813990.

Joergschmalenstroeer, Haeb-Umbach, R., 2010. Online Diarization of Streaming Audio-Visual Data for Smart Environments. IEEE Journal Of Selected Topics In Signal Processing, (4)5. 845-856. doi: 10.1109/JSTSP.2010.2050519.

Kang, H.S. et al., 2016. Smart manufacturing: past research, present findings, and future directions. Int. J. Precis. Eng. Manuf. Green Technol. 3 (1), 111–128.

Kantarci, B., Mouftah, H.T., 2014. Trustworthy sensing for public safety in cloud-centric internet of things. IEEE Int. Things J. 1 (4), 360–368.

Kaplan, S. et al., 2015. Driver behavior analysis for safe driving: a survey. IEEE Trans. Intell. Transport. Syst. 16 (6), 3017–3032. https://doi.org/10.1109/TITS.2015.2462084.

Kasteren, T. V., Krose, V., 2007. Bayesian Activity Recognition in Residence for Elders. in Proc. Third IET Int'l Conf. Intelligent Environments (IE '07), pp. 209-212.

Katre, P., et al., 2017. A Survey on Shortest path Algorithm for Road Network in Emergency Services. 2nd International Conference for Convergence in Technology (I2CT). pp. 393-396, doi: 10.1109/I2CT.2017.8226158.

Kaur, D. et al., 2018. Tensor-based big data management scheme for dimensionality reduction problem in smart grid systems: SDN Perspective. IEEE Trans. Knowl. Data Eng. 30 (10), 1985–1998. https://doi.org/10.1109/TKDE.2018.2809747.

Kesavaraj, G., Sukumaran, S., 2013. A study on classification techniques in data mining. in Proceedings of the 4th International Conference on Computing, Communications and Networking Technologies (ICCCNT '13), pp.1–7.

Knorr, E., Ng, R., 1998. Algorithm for mining distance-based outliers in large datasets. in proc. International Conference on Very Large Data Bases (VLDB'98), pp. 392-403.

Knorr, E., Ng, R., 1997. A unified notion of outliers: Properties and computation. in proc. 1997 International Conference on Knowledge Discovery and Data Mining (KDD'97), pp. 219-222.

Kravchenko, Y., et al., 2017. Technology Analysis for smart Home Implementation. in proc. of 4th International Scientific-Practical Conference on Problems of Infocommunications, Science and Technology. Kharkiv, Ukraine. Oct. 2017, pp. 579-584.

Kyaw, T. Y., et al., 2017. Smart Aquaponics System for urban farming. The 15th International Symposium on District Heating and Cool, World Engineers Summit – Applied Energy Symposium & Forum: Low Carbon Cities & Urban Energy Joint Conference, WES-CUE 2017, Singapore. (143), pp. 342-347.

Lai, Y.-X., Lai, C.-F., Huang, Y.-M., Chao, H.-C., 2013. Multi-appliance recognition system with hybrid SVM/GMM classifier in ubiquitous smart home. Inform. Sci. 230, 39–55.

Lanzisera, S. et al., 2014. Communicating power supplies: bringing the internet to the ubiquitous energy gateways of electronic devices. IEEE Internet Things J. 1 (2), 153–160.

Lee, H. S., Bang, H. C., 2013. Detecting counterfeit products using supply chain event mining. 15th International Conference on Advanced Communications Technology (ICACT), pp. 744–748.

Lee, J., et al., 2014. Service innovation and smart analytics for Industry 4.0 and big data Environment. Product Services Systems and Value Creation. Proceedings of the 6th CIRP Conference on Industrial Product-Service Systems, (16), pp. 3–8.

Lee, Y., Yang, W., Kwon, T., 2018. Data Transfusion: pairing wearable devices and its implication on security for internet of things. IEEE Access 6, 48994–49006. https://doi.org/10.1109/ACCESS.2018.2859046.

Leonardo de M. B., Dib, A., et al., 2017. Hybrid PLC/Wireless Communication For Smart Grids and Internet of Things Applications. IEEE Internet of Things. 2017. (5)2, 655-667. doi: 10.1109/JIOT.2017.2764747.

Li, L. et al., 2017a. Sequential behavior pattern discovery with frequent episode mining and wireless sensor network. IEEE Commun. Mag. 55 (6), 205–211.

Li, T. Y., et al., 2017. A Supporting System for Quick Dementia Screening Using PIR Motion Sensor in Smart Home. IEEE International Conference on Systems, Man, and Cybernetics (SMC) Banff Center, Banff, Canada. pp. 1369-1374, doi: 10.1109/SMC.2017.8122804.

Li, Y. et al., 2017c. Smart choice for the smart grid: narrowband internet of things (NB-iot). IEEE Int. Things J. 5 (3), 1505–1515. https://doi.org/10.1109/JIOT.2017.2781251.

Liu, B., 2007. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. Data-Centric Systems and Applications, Database Management & Information Retrieval, Computer Science, 2nd ed., Springer-Verlag Berlin Heidelberg, 2011. doi: 10.1007/978-3-642-19460-3.

Lyu, L. et al., 2018. PPFA: Privacy Preserving Fog-enabled Aggregation in Smart Grid. IEEE Trans. Ind. Inform. 14 (8), 3733–3744. https://doi.org/10.1109/TII.2018.2803782.

Mahmood, K. et al., 2017. An elliptic curve cryptography based lightweight authentication scheme for smart grid communication. Future Generat. Comput. Syst. 81, 557–565. https://doi.org/10.1016/j.future.2017.05.002.

Marah, R., Hibaoui, A. E., 2018. Algorithms for Smart Grid Management. Sustainable Cities and Society – Special issue: Microgrids Implementation and Optimization. (38), 627-635.

Marjani, M., et al., 2017. Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges. IEEE Access (5), 5247-5261.d oi: 10.1109/ACCESS.2017.2689040.

Miorandi, D., Sicari, S., De Pellegrini, F., et al., 2012. Internet of things: vision, applications and research challenges. Elsevier J. Ad Hoc Netw. 10 (7), 1497–1516.

Misgeld, B.J.E., Luken, M., Heitzmann, D., Wolf, S.I., Leonhardt, S., 2016. Body-sensor-network-based spasticity detection. IEEE J. Biomed. Health Inform. 20 (3), 748–755.

Moens, S., Aksehirli, E., Goethals, B., 2013. Frequent Itemset Mining for Big Data. Big Data, IEEE International Conference on, pp.111, 118.

Mohammadi, M., Al-Fuqaha, A., Sorour, S., Guizani, M., 2018. Deep learning for IoT big data and streaming analytics: a survey. IEEE Commun. Surveys Tutorials 20 (4), 2923–2960. https://doi.org/10.1109/COMST.2018.2844341.

Munshi, A.A. et al., 2017. Big data framework for analytics in smart grids. Electric Power Syst. Res. 151, 369–386. https://doi.org/10.1016/j.epsr.2017.06.006.

Nahar, S., Zhong, T., Monday, H. N., Mills, M. O., Nneji, G. U., Abubakar, H. S., 2019. A Survey on Data Stream Mining Towards the Internet of Things Application. 4th Technology Innovation Management and Engineering Science International Conference (TIMES-iCON), Bangkok, Thailand, pp. 1-5. doi: 10.1109/TIMES-iCON47539.2019.9024597.

Neto, A. et al., 2018. Fog-based crime-assistance in smart IoT transportation system. IEEE Access 6, 11101–11111. https://doi.org/10.1109/ACCESS.2018.2803439.

Oh, C., Hwang, D., Lee, T., 2015. Joint access control and resource allocation for concurrent and massive access of M2M Devices. IEEE Trans. Wireless Commun. 14 (8), 4182–4192. https://doi.org/10.1109/TWC.2015.2417873.

Ortiz, A.M., Hussein, D., Park, S., et al., 2014. The cluster between internet of things and social networks: review and research challenges. IEEE Internet Things J. 1 (3), 206–215.

Otuoze, A.O. et al., 2017. Smart grids security challenges: classification by sources of threats. J. Electr. Syst. Inf. Technol. 5 (3), 468–483.

Pacheco, F., Exposito, E., Gineste, M., Baudoin, C., Aguilar, J., 2019. Towards the deployment of machine learning solutions in network traffic classification: a systematic survey. IEEE Commun. Surv. Tutorials 21 (2), 1988–2014. https://doi.org/10.1109/COMST.2018.2883147.

Park, C., et al., 2017. Dynamic Characteristics of Smart Grid Technology Acceptance. International Scientific Conference on Environmental and Climate Technologies (CONECT17), Riga, Latvia. https://doi.org/10.1016/j.egypro.2017.09.040.

Park, J. S., et al., 1995. An effective hash-based algorithm for mining association rule. in proc. 1995 ACM-SIGMOD International Conference on Management of Data (SIGMOD'95), pp. 175–186.

Perera, K., Dias, D., 2011. An intelligent driver guidance tool using location based services. in Proc. International Conference on Spatial Data Mining and Geographical Knowledge Services, pp. 246–251.

Phuttharak, J., Loke, S.W., 2019. A review of mobile crowdsourcing architectures and challenges: toward crowd-empowered internet-of-things. IEEE Access 7, 304–324. https://doi.org/10.1109/ACCESS.2018.2885353.

Popovic, T. et al., 2017. Architecting an IoT-enabled platform for precision agriculture and ecological monitoring: a case study. Comput. Electron. Agric. Elsevier 140, 255–265.

Qiu, J., Tian, Z., Du, C., Zuo, Q., Su, S., Fang, B., 2019. A survey on access control in the age of internet of things. IEEE Internet Things J. 7 (6), 4682–4696. https://doi.org/10.1109/JIOT.2020.2969326.

Qolomany, B. et al., 2019. Leveraging machine learning and big data for smart buildings: a comprehensive survey. IEEE Access 7, 90316–90356. https://doi.org/10.1109/ACCESS.2019.2926642.

Rad, A.B., Eftestøl, T., Engan, K., Irusta, U., Kvaløy, J.T., Kramer-Johansen, J., Wik, L., Katsaggelos, A.K., 2017. ECG-based classification of resuscitation cardiac rhythms for retrospective data analysis. IEEE Trans. Biomed. Eng. 64 (10), 2411–2418. https://doi.org/10.1109/TBME.2017.2688380.

Rad, A. B., Eftestøl, T., Engan, K., Irusta, U., Kvaløy, J. T., Kramer-Johansen, J. Wik, L., Katsaggelos, A. K., 2014. Nearest-manifold classification approach for cardiac arrest rhythm interpretation during resuscitation. in Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference. pp. 3621-3625, doi: 10.1109/ICASSP.2014.6854276.

Rahim, M. H., et al., 2018. Energy Efficient Smart Building Using Coordination among Appliances Generating Large Data. in Proceeding IEEE Access. (6), 34670-34690. doi: 10.1109/ACCESS.2018.2805849.

Rashid, M.M., Kamruzzaman, J., Hassan, M.M., Shahriar Shafin, S., Bhuiyan, M.Z.A., 2020. A survey on behavioral pattern mining from sensor data in internet of things. IEEE Access 8, 33318–33341. https://doi.org/10.1109/ACCESS.2020.2974035.

Rashidi, P. et al., 2011. Discovering activities to recognize and track in a smart environment. IEEE Trans. Knowl. Data Eng. 23 (4), 527–539. https://doi.org/10.1109/TKDE.2010.148.

Razzaque, M.A., Milojevic-Jevric, M., Palade, A., et al., 2016. Middleware for internet of things: a survey. IEEE Internet Things J. 3 (1), 70–95.

Roddick, J.F. et al., 2002. A survey of temporal knowledge discovery paradigms and methods. IEEE Trans. Knowl. Data Eng. 14 (4), 750–767. https://doi.org/10.1109/TKDE.2002.1019212.

Rosslin, J., Tai-hoon, K., 2010. Application, systems and methods in smart home technology: a review. Int. J. Adv. Sci. Technol. 15, 37–48.
Saini, M., Alam, K.M., Guo, H., et al., 2017. InCloud: a cloud-based middleware for vehicular infotainment systems. Multimed Tools Appl. 76, 11621–11649. https://doi.org/10.1007/s11042-015-3158-4.
Saives, J. et al., 2015. Activity discovery and detection of behavioral deviations of an inhabitant from binary sensors. IEEE Trans. Automat. Sci. Eng. 12 (4), 1211–1224.
Samarah, S., et al., 2017. An Efficient Activity Recognition Framework: Toward Privacy-Sensitive Health Data Sensing. Special Section on Advances Of Multisensory Services And Technologies For Healthcare In Smart Cities, (5), pp. 3848-3859.
Schuh, G. et al., 2014. Collaboration Mechanisms to increase Productivity in the Context of Industrie 4.0. Robust Manufacturing Conference (RoMaC 2014), Elsevier, (19), pp. 51 – 56.
Severino, G., D'Urso, G., Scarfato, M., Toraldo, G., 2018. The IoT as a tool to combine the scheduling of the irrigation with the geostatistics of the soils. Future Gener. Comput. Syst. 82, 268–273. https://doi.org/10.1016/j.future.2017.12.058.
Shah, Z. et al., 2015. A spatio-temporal data summarization paradigm for real-time operation of smart grid. J. Lteax Class Files 14, 8. https://doi.org/10.1109/TBDATA.2017.2691350.
Sheikholeslami, G., et al., 1998. WebCluster: A multi resolution clustering approach for very large spatial databases. in proc. International Conference on Very Large Data Bases (VLDB'98), pp. 428-439.
Shu, L., Mukherjee, M., Pecht, M., Crespi, N., Han, S.N., 2018. Challenges and research issues of data management in IoT for large-scale petrochemical plants. IEEE Syst. J. 2 (3), 2509–2523. https://doi.org/10.1109/JSYST.2017.2700268.
Shukla, S., et al., 2017. Survey of Various Data Collection Ways for Smart Transportation Domain of Smart City. International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2017). pp. 681-685, doi: 10.1109/I-SMAC.2017.8058265.
Silva, D. D., 2016. A Data Fusion Technique for Smart Home Energy Management and Analysis. Industrial Electronics Society, IECON 2016 - 42nd Annual Conference of the IEEE, Florence, Italy, pp. 4594-4600, doi: 10.1109/IECON.2016.7793298.
Stankovic, J.A., 2014. Research directions for the internet of things. IEEE Internet Things J. 1 (1), 03–09.
Stojmenovic, I., 2014. Machine-to-machine communications with in-network data aggregation, processing, and actuation for large-scale cyber-physical systems. IEEE Internet Things J. 1 (2), 122–128.
Sushmitaruj, Nayak, A., 2013. A Decentralized Security Framework for Data Aggregation and Access Control in Smart Grids. IEEE Transactions on Smart Grid. (4)1, 196-205. doi: 10.1109/TSG.2012.2224389.
Talavera, J.M. et al., 2017. Review of IoT applications in agro-industrial and environmental fields. Comput. Electron. Agric. Elsevier 142, 283–297.
Tan, P.N., Kumar, V., Steinbach, M., 2006. Introduction to Data mining. Pearson Education publisher.
The State Council. The notice of the State Council on printing and distributing, 2015. Made in China 2025. [Internet]. May 8 [cited2016 Dec 20]. Available from: http://www.gov.cn/zhengce/content/2015-05/19/content_9784.htm. Chinese.
Thibaud, M., Chi, H., Zhou, W., Piramuthu, S., 2018. Internet of things (IoT) in high-risk environment, health and safety (EHS) industries: a comprehensive review. Decis. Support Syst. 108, 79–95. https://doi.org/10.1016/j.dss.2018.02.005.
Tian, Z., Gao, X., Su, S., Qiu, J., 2020a. Vcash: A novel reputation framework for identifying denial of traffic service in internet of connected vehicles. IEEE Internet Things J. 7 (5), 3901–3909. https://doi.org/10.1109/JIOT.2019.2951620.
Tian, Z., Luo, C., Qiu, X., Du, Guizani, M., 2020. A Distributed Deep Learning System for Web Attack Detection on Edge Devices. in IEEE Transactions on Industrial Informatics. (16)3, 1963-1971. doi: 10.1109/TII.2019.2938778.
Tsai C.W., et al., 2014. Data Mining for Internet of Things: A Survey. IEEE Communication Surveys and Tutorials. (16)1, 451.
Tu, C. et al., 2017. Big data issues in smart grid – A review. Renew. Sustain. Energy Rev. 79, 1099–1107.
Tuballa, M. L., Abundo, M. L., 2016. A review of the development of Smart Grid technologies. Renewable and Sustainable Energy Reviews, vol.59, pp. 710–725, Jan.
Tzounis, A. et al., 2017. Internet of Things in agriculture, recent advances and future challenges. Biosyst. Eng. ELSEVIER 164, 31–48.
U.S. Energy Information Administration, Annual Energy Outlook, 2018. Electricity supply, disposition, prices, and emissions, Reference case: 2016.Available: https://www.eia.gov/outlooks/aeo/pdf/AEO2018.pdf.
Uckelmann, D., Harrison, M., Michahelles, F., 2011. An architectural approach towards the future internet of things. Architecting the Internet of Things. 1st ed., 1–24.
United Nation. 2014. World Urbanization Prospect. [Online].Available: http://dl.acm.org/citation.cfm?Id=308574.308676.
Uusitola, M., 2006. Global vision for the future wireless world from the WWRF. IEEE Veh. Technol. Mag. 1 (2), 4–8.
Virone, G. et al., 2008. Behavioral patterns of older adults in assisted living. IEEE Trans. Inf. Technol. Biomed. 5, 3.
Wang, J. et al., 2018. Deep learning for smart manufacturing: Methods and applications. J. Manuf. Syst. 48, 144–156.
Wang, S., 2014. Application of high precision accuracy irrigation based on the fuzzy spatial data mining in 4G. Sixth International Conference on Intelligent Human-Machine Systems and Cybernetics. pp. 74-77, doi: 10.1109/IHMSC.2014.26.
Wang, W., Yang, J., Muntz, R.., 1997. STRING: A statistical information grid approach to special data mining. in proc. International Conference on Very Large Data Base (VLDB'97), pp. 186-195.
Wren, C., Munguia-Tapia, E., 2006. Toward Scalable Activity Recognition for Sensor Networks. in Proc. Workshop Location and Context-Awareness, pp. 218–235.
Wu, J., 2009. A Novel Artificial Neural Network Ensemble Model Based on K–Nearest Neighbor Nonparametric Estimation of Regression Function and Its Application for Rainfall Forecasting. Computational Sciences and Optimization, CSO 2009. International Joint Conference, pp. 44-48, doi: 10.1109/CSO.2009.307.
Xie, L., et al., 2008. Event Mining in Multimedia streams. Proceedings of the IEEE. (96)4, 23-35.
Xu, Yi, Helal, A., 2016. Scalable CLOUD-SENSOR ARCHITECTURE FOR THE INTERNET OF THINGs. IEEE Internet Things J. 3 (3), 285–298.
Yaghmaee, M.H. et al., 2018. A fog-based internet of energy architecture for transacive energy management systems. IEEE Internet Things J. 5 (2), 1055–1069. https://doi.org/10.1109/JIOT.2018.2805899.
Yahata, S., et al., 2017. A Hybrid Machine Learning Approach to Automatic Plant Phenotyping for Smart Agriculture. International Joint Conference on Neural Networks (IJCNN). pp. 1787-1793, doi: 10.1109/IJCNN.2017.7966067.
Yassine, A. et al., 2017. Mining human activity patterns from smart home big data for health care applications. IEEE Access 5, 13131–13141.
Youngblood M. G., Cook, D. J., 2007. Data Mining for Hierarchical Model Creation. IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications And Reviews, (37)4, 561-572. doi: 10.1109/TSMCC.2007.897341.
Younis, M., Fahmy, S., 2004. Distributed clustering in ad-hoc sensor networks: A hybrid, energy-efficient approach. in Proc. IEEE Infocom. pp. 640, doi: 10.1109/INFCOM.2004.1354534.
Yu, T., et al., 2017. Recursive Principal Component Analysis based Data Outlier Detection and Sensor Data Aggregation in IoT Systems. IEEE Internet of Things Journal, (4)6, 453.
Yuan, Z., et al., 2017. Smart Manufacturing for the Oil Refining and Petrochemical Industry. Engineering, ELSEVIER, Volume 3, Issue 2, Pages 179-182, April
Yue, H., Guo, L., Li, R., et al., 2014. DataClouds: enabling community-based data-centric services over the internet of things. IEEE Internet Things J. 1 (5), 472–482.
Zaki, M.., Parthasarathy, S., Ogihara, M., Li,W., "Parallel algorithms for discovery of association rules," Data Min. and Knowl. Disc., pp. 343–373.
Zdravevski, E., et al., 2017. Improving Activity Recognition Accuracy in Ambient-Assisted Living Systems by Automated Feature Engineering IEEE Access. (5), pp. 5262 – 5280.
Zhang, Q. et al., 2018. An adaptive dropout deep computation model for industrial iot big data learning with crowd sourcing to cloud computing. IEEE Trans. Ind. Informat. (Early Access). 15 (4), 2330–2337. https://doi.org/10.1109/TII.2018.2791424.
Zhang, T., Ramakrishnan, R., Livny, M., 1997. BIRCH: a new data clustering algorithm and its applications. Data Min. Knowl. Disc. 1 (2), 141–182.
Zhao, Q., Bhowmick, S. S., 2003. Sequential pattern mining: A survey. Technical Report, CAIS Nayang Technological University Singapore,Tech. Rep., pp. 1-27.
Zhu, L., et al., 2018. Big Data Analytics in Intelligent Transportation Systems: A Survey. IEEE Transactions on Intelligent Transportation Systems (Early Access), April. (20), 383-398. doi: 10.1109/TITS.2018.2815678.

Priyank Sunhare was born in Indore, India in 1989. He received the B.E. in Electronics and Communication Engineering from Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal in 2011 and M.Tech. degrees in Embedded Systems from the Devi Ahilya Vishwavidyalaya Indore, in 2013. He is currently pursuing the Ph.D. degree in Electronic and Communication Engineering from the Devi Ahilya Vishwavidyalaya Indore, India. From 2013 to 2015 he was an Assistant Professor with Electronics and Communication Engineering Department, Shri Vaishnav Institute of Technology and Science. Since 2015, he has been an Lecturer with the Electronics and Telecommunication Engineering Department, Government Polytechnic College, Dewas, India. His research interests include Real time computing, IoT Sensors and Actuators, Cloud base Embedded Computing, and Distributed Computing with Data Mining and Knowledge discovery algorithms.

Rameez Raja Chowdhary received the B. Sc., M. Sc., PhD degree in Electronics from the Devi Ahilya University, Indore, and the M.Tech. degree in Embedded Systems Technology from the SRM University, Chennai. His research interests include Network controlled robotics, Embedded Systems, and Real Time Systems. He Published papers in refereed international journals and conferences. Presently is a faculty member at Devi Ahilya University, Indore.

Manju K. Chattopadhyay completed her BSc Physics (Hons.) from Delhi University in 1998. She was a Gold medallist in M.Sc. (Electronics) in 2000 and qualified UGC-JRF-NET. Manju completed her Ph.D in Computer Science and Electronics in 2008 from Devi Ahilya University, Indore where she is a faculty member since 2006. Her research interests include III-V Semiconductor Device modelling, Internet of things, ARM Microcontroller based applications. She has more than 25 research publications in refereed international journals and conferences and is a reviewer with IEEE, Elsevier and Wiley's journals.