Accepted Manuscript

Big data for internet of things: A survey

Mouzhi Ge, Hind Bangui, Barbora Buhnova

PII: S0167-739X(17)31695-3
DOI: https://doi.org/10.1016/j.future.2018.04.053
Reference: FUTURE 4131

To appear in: *Future Generation Computer Systems*

Received date : 29 July 2017
Revised date : 4 April 2018
Accepted date : 19 April 2018

Big Data for Internet of Things: A Survey

Mouzhi Ge[a,b], Hind Bangui[a,b,c], Barbora Buhnova[a,b]

[a]*Institute of Computer Science, Masaryk University, Brno, Czech Republic*
[b]*Faculty of Informatics, Masaryk University, Brno, Czech Republic*
[c]*FSTG, Cadi AyyaD University, Marrakesh, Morocco*
{mouzhi.ge,hind.bangui,buhnova}@mail.muni.cz

Abstract

With the rapid development of the Internet of Things (IoT), Big Data technologies have emerged as a critical data analytics tool to bring the knowledge within IoT infrastructures to better meet the purpose of the IoT systems and support critical decision making. Although the topic of Big Data analytics itself is extensively researched, the disparity between IoT domains (such as healthcare, energy, transportation and others) has isolated the evolution of Big Data approaches in each domain. Thus, the mutual understanding across IoT domains can possibly advance the evolution of Big Data research in IoT.

In this work, we therefore conduct a survey on Big Data technologies in different IoT domains to facilitate and stimulate knowledge sharing across the IoT domains. Based on our review, this paper discusses the similarities and differences among Big Data technologies used in different IoT domains, suggests how certain Big Data technology used in one IoT domain can be re-used in another IoT domain, and develops a conceptual framework to outline the critical Big Data technologies across all the reviewed IoT domains.

Keywords: Big Data, data analytics, Internet of Things, healthcare, energy, transportation, building automation, Smart Cities.

☆Fully documented templates are available in the elsarticle package on CTAN.

1. Introduction

Internet of Things (IoT) is one of the most promising technologies in the current epoch. This research paradigm is characterized by using smart and self-configuring objects that can interact with each other via global network infrastructure. Therefore, these seamless interactions between large amounts of heterogeneous objects represent IoT as a disruptive technology that enables ubiquitous and pervasive computing applications [1]. Accordingly, a wide range of industrial IoT applications [2, 3, 4] have been developed and deployed in different domains such as transportation, agriculture, energy, healthcare, food processing industry, military, environmental monitoring, or security surveillance.

Since IoT connects the sensors and other devices to the Internet, it plays an important role to support the development of smart services. In other words, the dynamic things collect different kinds of data from the real-world environment. Afterwards, the extraction of relevant information from IoT data can be used to improve and enrich our daily life with context-aware applications, which can for example display contents related to the current situation of the user. Further, context can be defined as the information that is used to characterize the situation of entities (i.e. whether a person, place or object) and the situation is considered to be relevant to the real-time interaction between a user and an application, including the user and the application themselves [5]. AS context is typically featured by location, time, state of people, and environmental settings, IoT becomes an important source of contextual data with an enormous volume, variety and velocity, which makes it an interesting and challenging domain for Big Data research.

Big Data has been classified according to five fundamental elements, which are volume (size of data), variety (different types of data from several sources), velocity (data collected in real time), veracity (uncertainty of data) and value (benefits to various industrial and academic fields). Moreover, other research work like [6] introduces additional characteristics beyond the 5Vs model such as: validity (correct processing of the data), variability (context of data), viscosity (latency data transmission between the source and destination), virality (speed of the data sent and received from various sources) and visualization (interpretation of data and identification of the most relevant information for the users). Despite the existence of additional characteristics of Big Data, the 5V model lays the foundational description of the Big Data concept [7, 8]. Recently, Big Data research has been undergoing substantial transformation from its research harvest towards its high impact and applications in different areas.

The fusion of Big Data and IoT technologies has created opportunities for the development of services for many complex systems like Smart Cities. Several Big Data technologies have emerged to support the processing of large volumes of IoT data [104], which are collected from different sources in the smart environment. However, the advancement of IoT and its applications in many different domains are causing a significant increase of vast amount and different types of data [10]. At the same time, Big Data and its technologies have opened new application opportunities for industries and academia to develop new IoT solutions. Therefore, the fusion of Big Data and IoT, as well as the highly dynamic evolution of the two domains, create new research challenges, which however have so far not been recognized and addressed by the research community.

Many existing works recognize the research trends in Big Data and IoT respectively, for example, many surveys have been conducted recently on Big Data technologies [69, 62, 74, 9, 63], as well as on IoT technologies [66, 67, 68]. However, there is a lack of studies that would interlink the two research communities. This stems naturally from the diversity of different IoT domains, which implies that the research in each domain (whether healthcare, agriculture, smart city, military, energy, or other) is identifying the best fitting Big Data technologies in isolation from the state of the research in others [146, 75, 64, 105]. Thus it narrows the space for knowledge sharing among these domains whose data still share many unrecognized characteristics, such as being often the time series of small homogeneous sensor readings with questionable data quality or high dynamism of data-generating devices.

The aim of this paper is to create a platform for mutual understanding of similarities and differences in Big Data research in different IoT domains. To achieve this objective, we have performed a survey of 117 papers applying Big Data techniques in eight IoT domains (healthcare, energy, transportation, building automation, smart cities, agriculture, industry and military) and used the results of this survey to identify the similarities, differences and opportunities that can be drawn from deep understanding of the big picture. In order to facilitate the paper analysis, we have firstly derived four widely accepted stages in Big Data process. Then the Big Data approaches used in the IoT domains are structured along with the Big Data stages. As far as we know, this is the first work that addresses the Big Data research and technologies in different IoT domains.

The main contribution of this paper is fourfold: (1) analysis of the state-of-the-art Big Data research in IoT, (2) comparison of the Big Data technologies across IoT domains, suggesting which Big Data technologies can possibly be used in other IoT domains, (3) analysis and interpretation of Big Data research by each IoT domain such as the pros and cons of Big Data technologies and applications in certain IoT domain, and (4) a conceptual framework to guide researchers and practitioners to select the Big Data technologies that are prevalently used in certain IoT domain.

The remainder of the paper is organized as follows. Section 2 describes the IoT domains that have been used to classify the relevant Big Data approaches. Similarly, Section 3 focuses on Big Data processes and its life cycle to derive the categories that can serve the purpose of classifying Big Data technologies according to their scope and purpose. The key results of the paper can be found in Section 4 and Section 5, which carry out a literature review on Big Data works applied to different IoT contexts and a comparison of IoT Domains from the Big Data perspective. Together with the discussions of major findings, research challenges and opportunities are derived within the two sections. Finally, Section 6 concludes the work and outlines the future research.

2. IoT Domains

IoT technologies have been incorporated into various important domains in our life. Over the past years, many traditional domains such as manufacture industry, healthcare or energy have become IoT-based and gained the capability of communicating among machines and human, as well as production of enriched data. As a result, these smart things/objects facilitate the creation of a modern, smart and autonomous domain around the IoT concept, which is a prerequisite to successful IoT adoption.

IoT can be considered as a federation of application contexts such as healthcare or transportation that require the adjustment of techniques to make them better fit the needs of that very context. Therefore, IoT domains refer to the IoT techniques that are applied in certain context such as healthcare IoT or transportation IoT. Futhermore, different IoT domains share a set of common features. For example, most of the IoT domains emphasize the data collection, monitoring, sharing, automation, control and collaboration. Also, their datasets usually consist of relatively homogeneous data records e.g. from sensors and other IoT devices, which are often in a time series. Further, most domains need to be robust against unreliable or unavailable IoT objects and security threats implied by the extent of the networks (such as injected data or stolen data) [65, 66, 67, 68].

In the following, we describe the IoT domains where Big Data approaches are applied. In order to structure the IoT domains, we adopt the classification scheme by Bahga and Madisetti [65] and adapt it slightly by placing less emphasis on environment and retail because of their strong intersection with other domains, but adding military, which is emerging as a new and promising IoT domain. Not considering environment and retail as stand-alone domains is in line with other IoT surveys [68, 66, 67] and is motivated by the fact that existing works on environment often fall either within energy, agriculture or smart cities domain, and existing works on retail are typically classified as part of the industry domain.

Healthcare. The main purpose of applying IoT in healthcare is to gather and analyze real-time medical information in order to minimize the limitations of traditional medical treatment (i.e. medical errors) [11, 12]. Moreover, cloud platforms are used to store and analyze the collected medical data stream [13, 14]. Consequently, the gathered information about the patient's health status allows the healthcare organizations to develop ubiquitous healthcare applications and optimize the existing services and solutions, i.e. applications for remote monitoring, nutrition, medicinal products, medical devices, medical facility, or health insurance. Hence, the application of IoT in healthcare domain aids to find the best health condition and healing plan for patients [6].

Energy. Nowadays, energy is mostly featured by smart grid IoT, which is an emerging intelligent electricity distribution system that aims at integration of renewable resources in power systems, greater control of the grid for its operators and consumer engagement in optimal power consumption [16, 17]. Besides, smart grid offers many valuable services [18] such as distribution and consumption management, transmission, advanced metering infrastructure, renewable energy integration, self-healing systems and energy storage. For these reasons, the smart grid has been considered as one of the smart technologies that will contribute in the development of Smart Cities by guaranteeing the efficiency, reliability, sustainability and safety of the electric grid [19, 20, 21]. To achieve these goals, smart grid cooperates with IoT technologies to create various intelligent services [22], and it uses Big Data technologies to achieve the insightful intelligence and efficient power management quality [23].

Transportation. Due to IoT technologies, the intelligent transportation systems have become more ubiquitous [24, 25]. Since transportation is one of the key activities for each citizen, the IoT sensors produce each day a significant amount of data [26] that can be used to guide route planning and develop the applications for surveillance, emergency management, traffic control, anomaly detection, situation recognition and traffic prediction. In addition, the shared transportation data can minimize the risk of pollution and traffic accidents that might damage the health of citizens. Therefore, the information sharing through IoT devices also contributes to a sustainable smart environment [27].

Building automation. The integration of a large number of heterogeneous IoT devices installed in smart buildings, i.e. homes, faculties and offices, is enabling monitoring of everyday activities of the citizens as well as predicting their future actions [30, 31]. As a result, IoT devices in smart buildings gather sensitive information that is timely and describes very detailed interactions between humans and machines [32, 33, 34]. Hence, understanding how general and personal information can advance the smart building research and is critical to optimize the smart services offered in smart buildings, such as security, access control, digital video, intrusion detection, fire detection and alarm, indoor air quality services or lighting control.

Smart Cities. The vision of Smart Cities is to improve the lifestyle of citizens by providing smart applications in various fields. To achieve this goal, the city employs IoT technologies to optimize different public systems and services, such as car parking, city cleaning, waste management, street lightning and emergency control [35, 36, 37]. The big picture for Smart Cities can be illustrated with existing projects, such as SmartSantander [38], which is being considered as a practical system for a large-scale Smart City test bed. To keep up with the growth of SmartSantander, City Data and Analytics Platform (CiDAP) [39] has been created to process both historical and real-time data produced in this smart environment. These two and other similar projects demonstrate that the collection of data is a key initiative to get experience and knowledge necessary to support the rapid development of the Smart City [40, 41].

Agriculture. Agriculture is a vital domain of our society that also takes advantage of the benefits from IoT technologies to assure the quality of the products and the satisfaction of end-customers [42]. For example, monitoring from IoT devices plays an important role to protect the agricultural products from attacks by rodents or insects [43, 44]. To effectively manage all the agricultural activities and find the optimal environmental conditions, cloud platforms are used to store and analyze the sensed information and in turn improve the agricultural productivity as well as save energy [45, 46].

Industry. The development of IoT applications for future industrial automation is a highly promising topic in the industry and manufacture domain. In fact, modern industrial companies adopt the IoT research to boost the growth of the global economy and to keep competitive advantages [47, 48]. There are many IoT applications in industry such as the development of industrial IoT competences by integrating IoT technology and its usage for the training in digital learning factories [49], product supply chain management [50], Machine-to-Machine (M2M) communications [51] or energy saving [52, 53]. It is often focused on the deployment and exploitation of company's own or shared data to provide suitable goods that the customer favors as well as the improvement of the functionality of the industrial IoT systems, i.e. performance evaluation, intelligent M2M communications, simulation, modeling and industrial wireless networks [54, 55].

Military. The application of IoT is also extended to the military domain [56] and brings a significant and valuable source of information that could improve the intelligence of various military applications [57, 58, 59] such as military logistics, surveillance and military robots. Furthermore, the integration of IoT in military domain is expected to save lives of citizens by detecting harmful chemicals or biological weapons. For these reasons, the management and analysis of the shared information is necessary to make the right protective decisions, provide guidelines to perform the tasks as well as understand, in real time, the implications of these decisions. For example, the Internet of Battle Things [60] is one of the IoT applications that introduces future smart battlefields. In this smart environment, the intelligent things are communicating, acting, and collaborating with one another without neither the presence nor the coordination of human war-fighters [61].

3. Big Data Processes and Life Cycle

Big Data technologies include numerous activities, methods and techniques, each employed for slightly different purpose. To understand these techniques in the Big Data processing lifecycle, this section reviews existing works on the Big Data processes and distillates the employed activities that are later used to classify Big Data approaches applied in IoT.

The Big Data papers used for this purpose were selected by searching academic databases and well-known publishers such as Sciencedirect, Google Scholar, ACM Digital Library, IEEE Xplore Digital Library, Springer as well as general Google search with keywords like *Big Data Process* and *Big Data Lifecycle*. We limited the search to the up-to-date papers over the last 5 years, which is from 2013 to 2017. The search resulted in papers that contain a classification of a step-by-step Big Data process or Big Data lifecycle. Detailed descriptions for the whole process or lifecycle should exist in the papers. We paid special attention to the survey papers on Big Data research. We thus selected six papers, which are Gandomi and Haider 2015 [70], Paakkonen and Pakkala 2015 [72], Khan et al. 2014 [71], Tsai et al. 2015 [74], Chen et al. 2014 [69] and Siddiqa et al. 2016 [73].

Gandomi and Haider 2015 [70] have divided the Big Data process into two main stages: data management and data analytics. Data management is to collect and store the data as well as clean and retrieve the data for the analysis preparation. The other process is data analytics, which deals with extracting insights from the data. It involves modelling, analysis and interpretation. Paakkonen and Pakkala 2015 [72] developed a Big Data reference architecture that considers data source and data storage as the input and infrastructure for Big Data process. The Big Data process consists of five main phases, which are data extraction, data loading and pre-processing, data processing, data analysis, data loading and transformation, and data visualization. Khan et al. 2014 [71] proposed a lifecycle of Big Data. This proposed lifecycle covers raw data collection, data classification, data analysis, and data retrieval for decision making.

In the lifecycle, data storing, sharing and security are also discussed. Tsai et al. 2015 [74] considered the Big Data process from an input and output perspective. The input phase is more like the data management proposed by Gandomi and Haider 2015 [70], which includes the data preparation operations such as data gathering, selection and transformation for data analytics. Then data analytics is the connection between data input and output. The activities in data analytics are for example data mining and pattern learning. In the output stage, the data evaluation and interpretation are applied to present the data analysis results. Chen et al. 2014 [69] summarized the Big Data activities into three aspects: data generation and acquisition, data storage and Big data applications. In each aspect, the related concepts are discussed, for example, data generation and acquisition are detailed to data generation, collection, transportation and pre-processing. It is also similar to the input stage in Tsai et al. 2015 [74] and data management stage in Gandomi and Haider 2015 [70]. Siddiqa et al. 2016 [73] proposed a Big Data management flow, which begins from Big Data sources to final decision making. There are five main steps in this flow: network management, storage management, pre-processing, processing and prediction. Security is across all the steps through the management flow.

Considering the six papers, we found that there is no unified taxonomy agreed for Big Data process and lifecycle. A variety of taxonomies for Big Data Process and Lifecycle have been proposed, where different taxonomies contain different classification granularity. However, there are some similarities between the taxonomies and terminologies used for Big Data process and lifecycle. For example, usually a set of first-order Big Data processes is proposed, under each process there are several sub-processes. For example, Siddiqa et al. 2016 proposed three steps for the Big Data processing, which are data storage, pre-processing and processing. Under pre-processing, there are two sub-processes: transmission and cleansing. Furthermore, we found that in different taxonomies the terms are mixing. For example, one term may be not existing in the other taxonomy or one taxonomy uses one term as a first-order process, whereas the other taxonomy uses it as a sub-process. For example, Siddiqa et al. 2016 uses data pre-processing as a first-order process, while Chen et al. 2014 considers it as a sub-process of data generation and acquisition.

In order to derive a set of commonly agreed and critical terms for our survey in IoT, we selected the terms that appear the most across all Big Data processes and lifecycle taxonomies from the six literature, indicating the most overlapping alignment among the taxonomies. We select the important aspects that appear at least in half of the reviewed literature. In Table 1, it can be seen that four important aspects from Big Data processes and lifecycle are considered to be important across the selected literature: (1) Storage, (2) Cleaning/Cleansing, (3) Analysis/Analytics, and (4) Visualization. Furthermore, the four selected aspects are closely related to the characteristics of Big Data. Since Big Data can be fast-moving in large volume, data storage is thus dealing with data velocity and volume. Data cleaning covers the data variety and veracity, as data is usually generated from different sources with different quality levels. Data analytics and visualization consider the Big Data from the value-driven view and tend to offer insightful and valuable interpretations from the Big Data. Therefore, the four selected aspects can be considered as the representative stages for processing Big Data. When we apply the aspects to analyze the Big Data in IoT domains, it can be interesting to see which IoT domain focuses on which aspect to what extent. It also can indicate the challenging area of the Big Data research in the IoT domain. For example, Big Data visualization is not yet well developed in certain IoT domain.

Table 1: Important Big Data terms from the selected literature

	Gandomi and Haider 2015 [70]	Paakkonen and Pakkala 2015 [72]	Khan et al. 2014 [71]	Tsai et al. 2015 [74]	Chen et al. 2014 [69]	Siddiqa et al. 2016 [73]	Appearance Frequency
Acquisition	X				X		2
Aggregation	X						1
Annotation	X						1
Analytics	X	X	X	X	X		**5**
Classification						X	1

Cleaning or Cleansing	X	X	X			X	**4**
Compression		X					1
Clustering			X			X	2
Collection					X		1
Combining		X					1
Extraction	X	X					2
Evaluation				X			1
Filtering			X				1
Gathering				X			1
Generation					X		1
Indexing			X			X	2
Integration	X		X				2
Load		X					1
Modeling	X						1
Mining			X	X			2
Recording	X						1
Prediction						X	1
Representation	X		X				2
Replication		X				X	2
Retrieval			X				1
Stream processing		X					1
Searching			X				1
Selection				X			1
Storage			X		X	X	**3**
Transformation		X		X			2
Transportation					X	X	2
Visualization		X	X	X			**3**

In order to specify the selected Big Data aspects, we define the scope of the four selected aspects from Big Data processes and lifecycle as follows. Storage is dealing with how we store the data, for example, after collecting the sensor data from IoT devices, where and how the data are stored. Cleaning/Cleansing can be understood as a pre-processing phase before Big Data analytics, mainly used to integrate, check and improve the quality of the Big Data. Then the Analytics phase includes all the methods and models for data analysis and processing such as data clustering algorithms and MapReduce processing. Finally, Visualization is concerned with presentation and interpretation of data analytics results, which includes all the techniques related to data visualization. The scope of the four selected aspects is used to classify the keywords of the Big Data research in IoT literature in the next section.

4. Big Data approaches in different IoT domains

In this section, we focus on the analysis and classification of Big Data approaches applied in different IoT domains. The papers for this study were selected by searching academic databases listed in Section 3, with keywords characterizing the examined IoT domains, including their synonyms and variations (e.g. for transportation, the search included keywords like mobility, traffic management, logistics, route planning), used in combination with keywords characterizing the Big Data process activities, whether general (e.g. big data, data analytics) or specific (e.g. anomaly detection, data exploration, data observation, data summarization, data processing, data mining, machine learning, dataset, database, regression, aggregation/disaggregation, data visualization, data collection, data selection, data extraction, data integration, NoSQL search). We limited the search to the up-to-date papers over the last 5 years, which is from 2013 to 2017.

While the paper search is not limited to specific outlets, such as certain journals or conference proceedings, the academic disciplines of collected papers are scoped to computer science, information systems, computer networks and communications. In order to conduct an intercoder reliability check, each author has followed the search protocol and selected the papers independently. We have then synchronized the paper selection by the commonly agreed papers, which resulted in 139 papers in total. After collecting the search results, each paper underwent relevance check by all the authors, during which its relevance to both IoT and Big Data was verified.

The results are then summarized in a table form, as presented in the appendix. In Table 2, we show an excerpt from this table for the healthcare domain to demonstrate the strategy used to summarize the Big Data technologies used in this domain. Detailed discussion of the findings from all these tables follows in Section 5.

Table 2: Classification of Big Data approaches applied in Healthcare IoT

	Big Data			
	Storage	Cleaning/Cleansing	Analysis/Analytics	Visualization
Healthcare (35 papers)	Raspberry Pi [11], Cloud [13, 14, 15, 154, 153], HBase[6, 161], Spark SQL[6], Hive [6], Cloudera Impala [6], Apache Parquest [6], Python-based [157], Mysql [158], MongoDB Databases [158], NoSQL[159, 161], Centralized Database [165]	Integrity [15, 154], Information Exploration [161], Anomaly Detection [157, 158, 162, 164], Data Integration [159, 161], Reasoning [159], Coreference [159], Resolution [159], Entity Linking [159], Information Extraction [159, 165, 161], Consolidation [159], Paraphrase [159], Resolution [159], Ontology Alignment [159], Information Fusion [159, 161], Logical Inference [161], Data Transformation [161], Aggregation [161, 159, 163, 160]	Semantic Analysis [11], Clustering [161, 157, 159, 158, 162, 6, 164, 165, 163, 160, 13], AHP Method [162], Classification [161, 157, 159, 162, 160, 13, 14] Feature Extraction [154, 14, 155, 156], MapReduce [161, 6], Strom [6], Hadoop [161, 6], Apache Mahout [6], Spark [6], Association rule [158], Patterns [161, 158], Fuzzy [159], Regression [162], Descriptive Statistics [162], Deep Learning [161], Decision Trees [161] , Bayesian Fusion [161], Neural Network [161], Apache Pig [6], Apache Flume [6], Apache Hadoop Yarn [6], Semantic Mining [159], Knowledge Discovery [165]	Data Visualization [161, 157, 159, 6][162, 164, 165, 13] Interpretation [161, 157, 159] Data Lineage [15] [154] Visual Exploration [157], Visual Analytics [165]

As there might be no clear boundary between the IoT domains, especially for the Smart City, the concept can be enlarged to “Smart Everything”. Thus, if some papers span two or more domains, we assign the papers based on the paper’s main research focus. For papers related to Smart City, we differentiate the general Smart City paper and the narrower focus of an application context, for example, Smart transportation paper will be assigned to the transportation domain. The selected papers are firstly classified based on the IoT domains and

Big Data aspects, and then the papers are further summarized by the keywords, indicating which elements from Big Data are used in this IoT domain.

5. Comparison of IoT Domains from Big Data Perspective

This section summarizes the findings related to Big Data across IoT domains, as well as the findings related to IoT domains for Big Data technologies. Tables with all results supporting the discussed findings can be found in Appendix.

5.1. Findings related to Big Data across IoT domains

Storage. In all IoT domains, cloud storage has been the most widely accepted platform to store the Big IoT data. This is however not specific for the IoT data. It has been found that cloud storage is more suitable to store and scale the Big Data across different domains [194]. Upon the cloud storage, both NoSQL and Relational Database are used to store IoT data. For example, in the Smart Cities domain, IoT data is stored in NoSQL databases such as CouchDB [177] and MongoDB [177]. This can be because Smart City is a new IoT domain, thus it may have accepted more up-to-date storage technologies such as NoSQL for Big Data. On the other hand, some IoT domains like healthcare and agriculture are still using relational databases as storage. One possible reason can be that the IoT domains such as healthcare and agriculture are traditional domains. Although the applications in those traditional domains are dealing with huge amount of data, there might be legacy systems that are deployed in the relational databases. Also, we can observe that some IoT domain such as industry are using both NoSQL and relational databases. The industry IoT domain may be experiencing a transition period in the data storage, for example from Relational Database to NoSQL database, which can be more proper to scale the Big Data.

We further found that across all IoT domains, NoSQL database is more used than relational database in IoT applications. We consider that NoSQL database on the cloud platform is the favored solution for storing Big IoT Data. Therefore, for certain domains such as healthcare and agriculture, we suggest to move the storage of Big IoT Data to the NoSQL database. This may raise an issue of how to deal with the legacy storage. Thus some data conversion has been considered as one of the research topics. For example, with DB2RDF [132], the relational data can be transferred into the semantic triple store.

Cleaning/Cleansing. Among the IoT domains, we found that Big Data cleaning/cleansing includes two main sets of keywords, one set is regarding the data integration that intends to aggregate the IoT data from different sources. Since the IoT data can usually be located in different sites, most IoT domains have considered data integration as an important data preparation phase. In the data management research, data integration is usually used interchangeably with ETL (Extract, Transform, Load), which also can be observed in the Smart City domain. For example, while some papers [40] use ETL as the process of data integration, other papers such as [168] directly use the term of data integration. Furthermore, we have observed that in some of the reviewed IoT domains such as transportation, industry and agriculture, the concept of data integration is also termed as data fusion or data aggregation. Since many IoT data analytics or master data management initiatives are based on regular data integration, data integration has been considered as a prerequisite for further data analytics.

The other set of keywords is about the data quality management, which is used to deal with low-quality data such as corrupted data detection in the energy IoT, data redundancy reduction in building automation IoT and data integrity check in the healthcare IoT. Data quality research can be traced back to 80s [195]. With the advent of Big Data era, Big Data Quality has been emerging as a new research area [196]. In the IoT domains, the IoT data is featured by the Big Data 3Vs. For example, the sensor data from IoT is in large volume, generated from various sites and with high-speed updates such as streaming data. However, little research has been done towards the Big Data quality management in IoT domains. Notably, since the quality of the IoT data analytics is directly related to the quality of Big IoT Data, which is often poor due to cheap (and rather redundant) IoT devices, we suggest that the data analytics in IoT should be based on high-quality data. Big Data quality management is critical for the IoT applications, especially for the data analytics in IoT.

Analysis/Analytics. From our reviews, we found that there has been a variety of Big Data technologies that are used for data analytics in IoT domains. For example, some typical technologies such as Hadoop and Spark have been used in the healthcare and transportation domains. Thus in order to process the Big Data, MapReduce is a well accepted method in the IoT to perform parallel computing and distributed storage. As far as we found, there is no specific Big Data technologies designed for certain IoT domain. However, different algorithms are used to conduct data analytics in different IoT domains. For example, while feature extraction and decision trees are popular in the healthcare IoT, neural network and association rule mining are used in the energy IoT. Although the IoT domains are different, there is certain similarity at the level of IoT data types, such as all are coming from sensors. We therefore infer that some data analytics methods used in one domain can also be re-used in the other domain. As we have identified (see Appendix), the key methods for Big IoT data analytics are semantic analysis, analytic hierarchy process, clustering, feature extraction, association rule mining, pattern recognition, decision Tree, neural network, Bayesian network, frequent pattern mining, deep Learning, regression, fuzzy logic, rule extraction, genetic algorithm, multiple linear regression, Naive Bayes, K-nearest neighbor algorithm, contextual filtering, sequence analysis, and data envelopment analysis. From those candidate methods, certain IoT domain may choose or try to use some data analytics methods that are not yet applied in this domain.

Viewing the IoT data analytics methods, one of the most important and popular algorithms for data analytics across all IoT domains is clustering. The data from IoT devices are possibly generated from different data sources such as sensors and mobile devices. These Big IoT Data is in large volume, fast-moving and usually unstructured, in form of e.g. image data or stream data. The Big Data analytics mainly aims at firstly classifying the data, then mining the patterns and finally producing predictions. For example, in the transportation or Smart City IoT, there can be real-time traffic image data from various IoT devices such as road surveillance, satellite photos and traffic sensors. In order to analyze the real-time density of the cars, the traffic image data from different sources needs to be firstly clustered and then processed for further analysis. However, for data clustering, the road surveillance images data and stream traffic sensor data are considered as one IoT data input but with different data structures. From our review, we found that the creation of strategy for clustering the IoT data from different sources brings significant challenges in the IoT research.

Visualization. Regarding the data visualization in IoT domains, we found that there is a limited number of papers addressing the visualization of Big IoT Data. Among those papers we surveyed, some papers such as [15, 82, 121, 165] have mentioned visualization for Big IoT Data but visualization is not the main research focus in these papers. Across these IoT domains, the visualization of Big IoT Data is sometimes named as data interpretation or data presentation. There is a lack of specific Big Data visualization method for IoT that describes how to deal with pre-processing, processing and post-processing of visual data in real time. Moreover, the selected work that employed visual analysis algorithms usually neglect the utilization of machine learning or data mining to enhance the performance of visual analysis in terms of functionality, reliability, and scalability. Also, the integration of visual models with structured and semi-structured models is not well addressed in the IoT domains. Thus, from our review, we found that Big Data visualization methods are expected to be a promising challenge for future Big Data research in IoT.

Among the data visualization methods, visual analytics is the most used method for Big IoT Data visualization. It uses analytical reasoning from IoT data by interactive visual interfaces. For the IoT domains such as Smart Cities, industry and military, we found that visual analytics is not yet widely used. We suggest that it can be valuable to consider visual analytics as one of the visualization methods for the Big IoT Data.

5.2. Findings related to IoT domains for Big Data technologies

The Big Data research paradigm has affected all the IoT domains to ensure the sustainable development of the services provided to the end users. Since those domains used similar Big Data technologies to optimize their services, it is possible to coordinate the services between IoT domains, such as sharing the same deployment of Big Data applications for all the IoT domains. However, the exploitation of those technologies in IoT domains depends on the technical advancement of the IoT areas. From our review, we found that healthcare includes 25% of the selected papers. Energy papers have 17%, smart cities 13%, agriculture 9%, transportation 8%, industry 7%, military 6% and building automation 5%. Thus, the healthcare domain is a relatively mature domain that attracts many researchers. Also, due to the characteristics of each IoT domain, Big Data technologies have been used to ensure safety, reliability and efficiency of the IoT services. For instance, in the military domain, we found that little research contains Spark compared to the healthcare domain, which integrates different popular Big Data technologies. On the other hand, the requirements of IoT domains on Big Data tools are usually similar. For example, the importance of visualization technologies in Smart City papers is the same as in smart building papers. Moreover, some IoT domains adopt specific Big Data technologies with different popularity. For example, the exploitation of NoSQL in the selected papers is as follows: Healthcare 24%, energy 22%, transportation 9%, agriculture 7%, building automation 7%, industry 4% and military 7%, indicating that, the IoT domains have higher or lower percentage of the deployment of Big Data technologies. Different goals and challenges from each domain define the critical exploitation and selection of Big Data technologies.

Among the selected IoT domains, we found that Big Data research, especially Big Data analytics, is very popular in the healthcare IoT. One of reasons can be that the prediction from data analytics in healthcare is critical to support medical decisions. For example, Manogaran et al. in 2017 [6] proposed a Big Data based knowledge management system that is used to support the clinical decisions. Zhong et al. in 2017 [161] proposed a Big data architecture for enhancing the accuracy of healthcare predictions, in which revised fusion node and deep learning algorithms are described in detail. It can be observed that the healthcare domain provides more mature studies in Big Data notably for data analytics because the precision of healthcare predications can directly influence disease diagnosis or even people lives.

Another emerging IoT domain is the military IoT. Although this is not one of the most cited domains, we found that Big Data based IoT technologies has revolutionized the military sector [60]. The military IoT applications maintain and combine three strong interactions: machine to machine, machine to human (e.g. war-fighters) and human to human. Further, the contextual information is gathered not just from static sources like the traditional IoT-big Data treatments, but from dynamic and ubiquitous devices. We consider that complicated interactions and dynamic data sources will bring the next complex level for Big Data research. Although due to confidential reasons the selected papers in this domain provided limited detailed descriptions on how to use Big Data analytics tools to perform certain specific military tasks, the implications from the military IoT applications are expected to increase radically the development of Big Data in the other IoT domains.

Recently, the Smart City is emerging as a popular research topic. There has been a rapid increase in literature and projects about Big Data and IoT in Smart Cities. From the selected papers, Smart City IoT contains the highest number of papers on Big Data topic, indicating that there is a strong relation between Big Data research and Smart City IoT. Big Data and IoT are not only considered as two typical features in the Smart City. The interaction between the two elements such as analytics on the Big IoT Data in Smart Cities has also attracted significant research attention. We therefore suggest that other domains such as industry IoT, where there are not yet many research works on

Big Data, may refer to the Big Data approaches and models in the Smart City IoT, as Smart City IoT may have a higher level of maturity on the Big Data research.

By observing the Big Data technologies across different IoT domains, we have developed a conceptual framework that illustrates the current dominant Big Data technology in this domain and the widely used technology across different IoT domains for certain stage in the Big Data process. The result is thus presented in the form of two keywords "dominant Big Data technology in this domain / widely used technology across different IoT domains". If the two keywords are identical, it means that the dominant Big Data technology in this IoT domain is also the widely used technology across all other IoT domains, we only present this keyword once to represent the two cases. If there is no dominant technology in the Big Data processing stage for certain IoT domain, we use "-" to illustrate this result. This conceptual framework is described in Table 3.

Table 3: Conceptual framework of Big Data approaches applied in IoT

Big Data				
	Storage	Cleaning, Cleansing	Analysis, Analytics	Visualization
Healthcare	Cloud	Anomaly Detection/Data Integration	Clustering	Data Visualization
Energy	Cloud	Outlier Detection/Data Integration	Clustering	Data Visualization
Transportation	Cloud	Data collection/Data Integration	Clustering	Data Visualization
Building automation	Cloud	Data aggregation/Data Integration	Clustering	Data Visualization
Smart Cities	Cloud	Anomaly Detection / Data Integration	Clustering	Data Visualization
Agriculture	Cloud	Data Integration	Classification / Clustering	Data Visualization

Big Data				
	Storage	Cleaning, Cleansing	Analysis, Analytics	Visualization
Industry	-/Cloud	Data Integration	-/Clustering	Data Presentation/Data Visualization
Military	Cloud	Data Quality/Data Integration	Data filtering/Clustering	-/Data Visualization

Using this conceptual Framework, we can identify which Big Data technology is widely used in which IoT domain and the overview of widespread Big Data technology from all the domains. For example, in the data cleaning and cleansing stage, it can been seen that outliers detection is the dominant keyword, and from the observations in other IoT domains, data integration can also be important in the energy IoT domain. Thus this framework offers the similarity and differences among the Big Data technologies in different IoT domains. We further found that some keywords may share an overlap in the meaning. For example, in the data cleaning and cleansing stage, building automation considers data aggregation as the most popular Big Data technology, while data integration is the a widely used term across all the IoT domains. Data aggregation and data integration may intend to solve the same issue but solutions may differ. Therefore, even if the keywords are similar, it can be valuable to investigate how the technologies are related in practice.

6. Conclusion

In this paper, we have conducted an extensive survey on the papers bridging the IoT and Big Data communities. Based on the review of IoT papers, we have selected a set of typical IoT domains and described the features in each domain. By reviewing the Big Data papers, we have derived four important aspects from the Big Data process analysis. After specifying the IoT domains and important Big Data aspects, we have focused on the literature review across the IoT and Big Data research. The related papers have been classified based on the IoT domain that the paper belongs to, which Big Data techniques are used in this IoT domain and how Big Data techniques are applied in certain IoT domain. Furthermore, we have extracted, summarized and organized the keywords from reviewed papers to indicate which Big Data elements are used in each IoT domain.

As such, we are able to conclude mutual understandings and interpretations across IoT domains in terms of Big Data, whereby suggest which state-of-the-art Big Data technologies can be possibly used for which IoT domain, as well as outline the Big Data research opportunities in different IoT domains. We have further analyzed the Big Data research across IoT domains, for example, the pros and cons of Big Data technologies and their applications in certain IoT domain. Based on these results, we have proposed a conceptual framework to guide researchers and practitioners to select the Big Data technologies that are prevalently used in certain IoT domain. This also provides practical implications on how we can possibly improve or complement certain IoT domain by using the widely accepted Big Data technologies from other domains.

As future work, we plan to further investigate each of the IoT domains and understand the distinct features and functions in each domain, and how Big Data technologies are applied together with the domain features and functions. This can provide more detailed insights on why certain Big Data technologies are used in one IoT domain but not in another. Furthermore, we would like to build deeper understanding of IoT data sets that are characteristic in different IoT domains to further facilitate knowledge transfer among the domains.

One of the limitations of this work is that the selection of the keywords is manually summarized from our understanding and expertise. We have narrowed this limitation by scrutinizing the related papers, conducting the synchronization among the author's selections and classifications. Furthermore, the validity of our proposed table is also reviewed and confirmed by different Big Data experts.

Acknowledgements

The work was supported from European Regional Development Fund Project *CERIT Scientific Cloud* (No. CZ.02.1.01/0.0/0.0/16_013/0001802).

References

[1] Van Kranenburg, R., A critique of ambient technology and the all-seeing network of RFID, Publisher: Institute of Network Cultures, 2008.

[2] Da Xu, Li, Wu He, and Shancang Li, Internet of things in industries: A survey. IEEE Transactions on Industrial Informatics 10(4), pp. 2233-2243, 2014.

[3] Li, S., Tryfonas, T. and Li, H., The internet of things: a security point of view. Internet Research, 26(2), pp.337-359, 2016.

[4] Yuehong YI, Zeng Y, Chen X, Fan Y. , The internet of things in healthcare: an overview. Journal of Industrial Information Integration. 31(1), pp.3-13, 2016.

[5] Dey, Anind K., Understanding and using context, Personal and Ubiquitous Computing 5(1) pp. 4-7, 2001.

[6] Manogaran, Gunasekaran, Chandu Thota, Daphne Lopez, V. Vijayakumar, Kaja M. Abbas, and Revathi Sundarsekar., Big data knowledge system in healthcare. In Internet of Things and Big Data Technologies for Next Generation Healthcare, pp. 133-157. Springer International Publishing, 2017.

[7] Erl, T., Khattak, W. and Buhler, P., Big data fundamentals: concepts, drivers and techniques. Prentice Hall Press, 2016.

[8] Kitchin, Rob. Big dataHype or revolution., The SAGE handbook of social media research methods pp. 27-39, 2017.

[9] Chen, Yong, Hong Chen, Anjee Gorkhali, Yang Lu, Yiqian Ma, and Ling Li., Big data analytics and big data science: a survey. Journal of Management Analytics 3(1) pp. 1-42, 2016.

[10] Mavromoustakis, Constandinos X., George Mastorakis, and Jordi Mongay Batalla., Internet of Things in 5G Mobile Technologies. Modeling and Optimization in Science and Technologies. Springer International Publishing, 2016.

[11] Zgheib, Rita, Emmanuel Conchon, and Rmi Bastide., Engineering IoT Healthcare Applications: Towards a Semantic Data Driven Sustainable Architecture. In eHealth 360, pp. 407-418. Springer International Publishing, 2017.

[12] Qiu, Tie, Xize Liu, Min Han, Mingchu Li, and Yushuang Zhang, SRTS: A Self-Recoverable Time Synchronization for Sensor Networks of Healthcare IoT, Computer Networks, In Press, 2017.

[13] Abawajy, Jemal H., and Mohammad Mehedi Hassan, Federated Internet of Things and Cloud Computing Pervasive Patient Health Monitoring System. IEEE Communications Magazine 55(1) pp. 48-53, 2017.

[14] Hossain, M. Shamim, and Ghulam Muhammad, Cloud-assisted industrial internet of things (iiot)enabled framework for health monitoring. Computer Networks, 101, pp. 192-202, 2016.

[15] Manogaran, Gunasekaran, Chandu Thota, Daphne Lopez, and Revathi Sundarasekar, Big data security intelligence for healthcare industry 4.0, In Cybersecurity for Industry 4.0, pp. 103-126. Springer International Publishing, 2017.

[16] Wong, Steven, and Jean-Paul Pinard, Opportunities for smart electric thermal storage on electric grids with renewable energy. IEEE Transactions on Smart Grid 8(2), pp. 1014-1022, 2017.

[17] Chren, Stanislav, and Barbora Buhnova. Local load optimization in smart grids with Bayesian networks. In Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on, pp. 004021-004027. IEEE, 2016.

[18] Emmanuel, Michael, and Ramesh Rayudu, Evolution of dispatchable photovoltaic system integration with the electric power network for smart grid applications: A review. Renewable and Sustainable Energy Reviews, 67, pp. 207-224, 2017.

[19] Kolokotsa, Dionysia, The role of smart grids in the building sector. Energy and Buildings 116, pp. 703-708, 2016.

[20] Kadlec, Miroslav, Barbora Buhnova, Jan Tomsik, Jan Herman, and Katerina Druzbikova. Weather forecast based scheduling for demand response optimization in smart grids. In Smart City Symposium Prague (SCSP), pp. 1-6. IEEE, 2017.

[21] Rossi, Bruno, Stanislav Chren, Barbora Buhnova, and Tomas Pitner. Anomaly detection in Smart Grid data: An experience report. In Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on, pp. 002313-002318. IEEE, 2016.

[22] Collier, Steven E., The emerging enernet: Convergence of the smart grid with the internet of things. IEEE Industry Applications Magazine 23(2), pp. 12-16, 2017.

[23] Jaradat, Manar, Moath Jarrah, Abdelkader Bousselham, Yaser Jararweh, and Mahmoud Al-Ayyoub, The internet of energy: Smart sensor networks and big data management for smart grid. Procedia Computer Science 56, pp. 592-597, 2015.

[24] Cheng, Jialang, Weigang Wu, Jiannong Cao, and Keqin Li, Fuzzy Group-Based Intersection Control via Vehicular Networks for Smart Transportations. IEEE Transactions on Industrial Informatics, 13(2), pp. 751-758, 2017.

[25] He, Wu, Gongjun Yan, and Li Da Xu, Developing vehicular data cloud services in the IoT environment. IEEE Transactions on Industrial Informatics, 10(2), pp. 1587-1595, 2014.

[26] Ma, Xiaolei, Haiyang Yu, Yunpeng Wang, and Yinhai Wang, Large-scale transportation network congestion evolution prediction using deep learning theory. PloS one 10(3), pp. 1-17, 2015.

[27] Kyriazis, D., Varvarigou, T., White, D., Rossi, A. and Cooper, J., Sustainable smart city IoT applications: Heat and electricity management and Eco-conscious cruise control for public transportation, 14th IEEE International Symposium and Workshops on a World of Wireless, Mobile and Multimedia Networks, pp.1-5, 2013.

[28] Wang, Chao, Xi Li, Xuehai Zhou, Aili Wang, and Nadia Nedjah, Soft computing in big data intelligent transportation systems. Applied Soft Computing 38, pp. 1099-1108, 2016.

[29] Fumeo, Emanuele, Luca Oneto, and Davide Anguita, Condition based maintenance in railway transportation systems based on big data streaming analysis. Procedia Computer Science, 53, pp. 437-446, 2015.

[30] Stojkoska, Biljana L. Risteska, and Kire V. Trivodaliev, A review of Internet of Things for smart home: Challenges and solutions. Journal of Cleaner Production 140, pp. 1454-1464, 2017.

[31] Moreno, M. Victoria, Miguel A. Zamora, and Antonio F. Skarmeta, An IoT based framework for usercentric smart building services. International Journal of Web and Grid Services, 11(1), pp. 78-101, 2015.

[32] ul Hasan, Najam, Waleed Ejaz, Imran Baig, Manaf Zghaibeh, and Alagan Anpalagan, QoS-aware channel assignment for IoT-enabled smart building in 5G systems. 8th International Conference on Ubiquitous and Future Networks pp. 924-928, 2016.

[33] Dutta, Joy, and Sarbani Roy, IoT-fog-cloud based architecture for smart city: Prototype of a smart building. 7th International Conference on Cloud Computing, Data Science and Engineering-Confluence, pp. 237-242. IEEE, 2017.

[34] Gesvindr, David, Jana Michalkova, and Barbora Buhnova. System for Collection and Processing of Smart Home Sensor Data. In Software Architecture Workshops (ICSAW), 2017 IEEE International Conference on, pp. 247-250. IEEE, 2017.

[35] Talari, Saber, Miadreza Shafie-khah, Pierluigi Siano, Vincenzo Loia, Aurelio Tommasetti, and Joo PS Catalo, A review of smart cities based on the internet of things concept, Energies, 10(4), 2017.

[36] Centenaro, Marco, Lorenzo Vangelista, Andrea Zanella, and Michele Zorzi, Long-range communications in unlicensed bands: The rising stars in the IoT and smart city scenarios. IEEE Wireless Communications, 23(5), pp. 60-67, 2016.

[37] Rathore, M. Mazhar, Awais Ahmad, Anand Paul, and Seungmin Rho, Urban planning and building smart cities based on the Internet of things using big data analytics. Computer Networks, 101, pp. 63-80, 2016.

[38] Sanchez, Luis, Luis Muoz, Jose Antonio Galache, Pablo Sotres, Juan R. Santana, Veronica Gutierrez, Rajiv Ramdhany, SmartSantander: IoT experimentation over a smart city testbed. Computer Networks, 61, pp. 217-238, 2014.

[39] Cheng, Bin, Salvatore Longo, Flavio Cirillo, Martin Bauer, and Ernoe Kovacs, Building a big data platform for smart cities: Experience and lessons from santander. IEEE International Congress on Big Data, 2pp. 592-599. IEEE, 2015.

[40] Gomes, Eliza, Mario AR Dantas, Douglas DJ de Macedo, Carlos De Rolt, Marcelo Luiz Brocardo, and Luca Foschini, Towards an Infrastructure to

Support Big Data for a Smart City Project. 25th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE),pp. 107-112, IEEE, 2016.

[41] Moreno, M. Victoria, Fernando Terroso-Senz, Aurora Gonzlez-Vidal, Mercedes Valds-Vela, Antonio F. Skarmeta, Miguel A. Zamora, and Victor Chang, Applicability of big data techniques to smart cities deployments. IEEE Transactions on Industrial Informatics 13(2), pp. 800-809, 2017.

[42] Gardaevi, Gordana, Mladen Veleti, Neboja Maleti, Dragan Vasiljevi, Igor Radusinovi, Slavica Tomovi, and Milutin Radonji, The IoT Architectural Framework, Design Issues and Application Domains. Wireless Personal Communications 92(1), pp. 127-148, 2017.

[43] Baranwal, Tanmay, and Pushpendra Kumar Pateriya, Development of IoT based smart security and monitoring devices for agriculture. 6th International Conference on Cloud System and Big Data Engineering (Confluence), pp. 597-602. IEEE, 2016.

[44] Brun-Laguna, Keoma, Ana Laura Diedrichs, Javier Emilio Chaar, Diego Dujovne, Juan Carlos Taffernaberry, Gustavo Mercado, and Thomas Watteyne, A Demo of the PEACH IoT-based frost event prediction system for precision agriculture. 13th Annual IEEE International Conference on Sensing, Communication, and Networking, pp. 1-3, IEEE, 2016.

[45] Botta, Alessio, Walter De Donato, Valerio Persico, and Antonio Pescap, Integration of cloud computing and internet of things: a survey. Future Generation Computer Systems, 56, pp. 684-700, 2016.

[46] Jiang, Shuwen, Tianen Chen, and Jing Dong, Application and Implementation of Private Cloud in Agriculture Sensory Data Platform. In Computer and Computing Technologies in Agriculture IX: 9th IFIP WG 5.14 International Conference, Beijing, China. pp. 60-67. Springer International Publishing, 2016.

[47] Palattella, Maria Rita, Nicola Accettura, Luigi Alfredo Grieco, Gennaro Boggia, Mischa Dohler, and Thomas Engel, On optimal scheduling in duty-cycled industrial IoT applications using IEEE802. 15.4 e TSCH. IEEE Sensors Journal, 13(10), pp. 3655-3666, 2013.

[48] Lee, In, and Kyoochun Lee, The Internet of Things (IoT): Applications, investments, and challenges for enterprises, Business Horizons, 58(4), pp. 431-440, 2015.

[49] Gronau, Norbert, Andr Ullrich, and Malte Teichmann, Development of the Industrial IoT Competences in the Areas of Organization, Process, and Interaction Based on the Learning Factory Concept, Procedia Manufacturing 9, pp. 254-261, 2017.

[50] Jayaram, Athul, Lean six sigma approach for global supply chain management using industry 4.0 and IIoT, 2nd International Conference on Contemporary Computing and Informatics, pp. 89-94, IEEE, 2016.

[51] Meng, Zhaozong, Zhipeng Wu, Cahyo Muvianto, and John Gray. A Data-Oriented M2M Messaging Mechanism for Industrial IoT Applications. IEEE Internet of Things Journal, 4(1), pp. 236-246, 2017.

[52] Wang, Kun, Yihui Wang, Yanfei Sun, Song Guo, and Jinsong Wu, Green industrial Internet of Things architecture: An energy-efficient perspective, IEEE Communications Magazine, 54(12), pp. 48-54, 2016.

[53] Yang, Lijun, Chao Ding, Meng Wu, and Kun Wang, Robust Detection of False Data Injection Attacks for the Data Aggregation in Internet of Things based Environmental Surveillance, Computer Networks, 2017.

[54] Nobre, Gustavo Cattelan, and Elaine Tavares, Scientific literature analysis on big data and internet of things applications on circular economy: a bibliometric study, Scientometrics 111(1), pp. 463-492, 2017.

[55] Bilal, Muhammad, Lukumon O. Oyedele, Junaid Qadir, Kamran Munir, Saheed O. Ajayi, Olugbenga O. Akinade, Hakeem A. Owolabi, Hafiz A.

Alaka, and Maruf Pasha, Big Data in the construction industry: A review of present status, opportunities, and future trends, Advanced Engineering Informatics, 30(3), pp. 500-521, 2016.

[56] Raglin, Adrienne, Somiya Metu, Stephen Russell, and Peter Budulas. Implementing Internet of Things in a military command and control environment, In SPIE Defense+ Security. International Society for Optics and Photonics, 2017.

[57] Fraga-Lamas, Paula, Tiago M. Fernndez-Carams, Manuel Surez-Albela, Luis Castedo, and Miguel Gonzlez-Lpez, A review on internet of things for defense and public safety, Sensors 16(10), 2016.

[58] Tortonesi, Mauro, Alessandro Morelli, M. Govoni, James Michaelis, Niranjan Suri, Cesare Stefanelli, and S. Russell, Leveraging Internet of Things within the military network environment - Challenges and solutions. 3rd World Forum on Internet of Things, pp. 111-116, IEEE, 2016.

[59] Pradhan, Manas, Fahrettin Gkgz, Nico Bau, and Daniel Ota, Approach towards application of commercial off-the-shelf Internet of Things devices in the military domain, 3rd World Forum on Internet of Things (WF-IoT), pp. 245-250, IEEE, 2016.

[60] Kott, Alexander, Ananthram Swami, and Bruce J. West, The Internet of Battle Things, Computer 49(12), pp. 70-75, 2016.

[61] Grieco, Luigi Alfredo, Alessandro Rizzo, Simona Colucci, Sabrina Sicari, Giuseppe Piro, Donato Di Paola, and Gennaro Boggia. IoT-aided robotics applications: Technological implications, target domains and open issues, Computer Communications, 54, pp. 32-47, 2014.

[62] Sagiroglu, Seref, and Duygu Sinanc, Big data: A review, International Conference on Collaboration Technologies and Systems, IEEE, 2013.

[63] Bajaber, Fuad, Big data 2.0 processing systems: Taxonomy and open challenges, Journal of Grid Computing, 14(3), pp. 379-405, 2016.

[64] Hashem, Ibrahim Abaker Targio, The role of big data in smart city, International Journal of Information Management, 36(5), pp. 748-758, 2016.

[65] Arshdeep Bahga and Vijay Madisetti: Internet of Things (A Hands-on-Approach), 446 pages. VPT; 1 edition. ISBN 0996025510, 2014.

[66] Al-Fuqaha, Ala, Mohsen Guizani, Mehdi Mohammadi, Mohammed Aledhari, and Moussa Ayyash., Internet of things: A survey on enabling technologies, protocols, and applications, IEEE Communications Surveys & Tutorials 17(4), pp.2347-2376, 2015.

[67] Chen, Shanzhi, Hui Xu, Dake Liu, Bo Hu, and Hucheng Wang, A vision of IoT: Applications, challenges, and opportunities with china perspective, IEEE Internet of Things journal, 1(4), pp. 349-359, 2014.

[68] Yaqoob, Ibrar, Ejaz Ahmed, Ibrahim Abaker Targio Hashem, Abdelmuttlib Ibrahim Abdalla Ahmed, Abdullah Gani, Muhammad Imran, and Mohsen Guizani, Internet of Things Architecture: Recent Advances, Taxonomy, Requirements, and Open Challenges, IEEE Wireless Communications, 24(3), pp. 10-16, 2017.

[69] Chen, M., Mao, S. and Liu, Y., Big data: A survey. Mobile Networks and Applications, 19(2), pp. 171-209, 2014.

[70] Gandomi, A. and Haider, M., Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), pp.137-144, 2015.

[71] Khan, Nawsher, Ibrar Yaqoob, Ibrahim Abaker Targio Hashem, Zakira Inayat, Waleed Kamaleldin Mahmoud Ali, Muhammad Alam, Muhammad Shiraz, and Abdullah Gani, Big data: survey, technologies, opportunities, and challenges, The Scientific World Journal, 2014.

[72] Pkknen, Pekka, and Daniel Pakkala, Reference architecture and classification of technologies, products and services for big data systems, Big Data Research, 2(4), pp. 166-186, 2015.

[73] Siddiqa, Aisha, Ibrahim Abaker Targio Hashem, Ibrar Yaqoob, Mohsen Marjani, Shahabuddin Shamshirband, Abdullah Gani, and Fariza Nasaruddin, A survey of big data management: Taxonomy and state-of-the-art, Journal of Network and Computer Applications, 71, pp. 151-166, 2016.

[74] Tsai, Chun-Wei, Chin-Feng Lai, Han-Chieh Chao, and Athanasios V. Vasilakos, Big data analytics: a survey, Journal of Big Data, 2(1), 2015.

[75] Diamantoulakis, Panagiotis D., Vasileios M. Kapinas, and George K. Karagiannidis, Big data analytics for dynamic energy management in smart grids, Big Data Research, 2(3), pp. 94-101, 2015.

[76] Chen, Wen, Kaile Zhou, Shanlin Yang, and Cheng Wu, Data quality of electricity consumption data in a smart grid environment, Renewable and Sustainable Energy Reviews, 75, pp. 98-105, 2017.

[77] Neagu, Bogdan Constantin, Gheorghe Grigora, and Florina Scarlatache, Outliers discovery from Smart Meters data using a statistical based data mining approach, 10th International Symposium on Advanced Topics in Electrical Engineering, pp. 555-558. IEEE, 2017.

[78] Rusitschka, Sebnem, Kolja Eger, and Christoph Gerdes, Smart grid data cloud: A model for utilizing cloud computing in the smart grid domain, First IEEE International Conference on Smart Grid Communications, pp. 483-488. IEEE, 2010.

[79] Baek, Joonsang, Quang Hieu Vu, Joseph K. Liu, Xinyi Huang, and Yang Xiang, A secure cloud computing based framework for big data information management of smart grid, IEEE transactions on Cloud Computing, 3(2), pp. 233-244, 2015.

[80] Sancho-Asensio, Andreu, Joan Navarro, Itziar Arrieta-Salinas, Jos Enrique Armendriz-igo, Virginia Jimnez-Ruano, Agustn Zaballos, and Elisabet Golobardes, Improving data partition schemes in Smart Grids via

clustering data streams, Expert Systems with Applications, 41(13), pp. 5832-5842, 2014.

[81] Aung, Zeyar, Database systems for the smart grid, In Smart Grids, pp. 151-168. Springer London, 2013.

[82] Tu, Chunming, Xi He, Zhikang Shuai, and Fei Jiang, Big data issues in smart grid - A review, Renewable and Sustainable Energy Reviews, 79, pp. 1099-1107, 2017.

[83] Aiello, Marco, and Giuliano Andrea Pagani, The smart grid's data generating potentials, Federated Conference on Computer Science and Information Systems, pp. 9-16. IEEE, 2014.

[84] Biscarri, Flix, Iigo Monedero, Antonio Garca, Juan Ignacio Guerrero, and Carlos Len., Electricity clustering framework for automatic classification of customer loads, Expert Systems with Applications, 2017.

[85] Daki, Houda, Asmaa El Hannani, Abdelhak Aqqal, Abdelfattah Haidine, and Aziz Dahbi, Big Data management in smart grid: concepts, requirements and implementation, Journal of Big Data, 4(1), 2017.

[86] Zhou, Dao, Jiahui Guo, Ye Zhang, Jidong Chai, Hesen Liu, Yong Liu, Can Huang, Xun Gui, and Yilu Liu, Distributed data analytics platform for wide-area synchrophasor measurement systems, IEEE Transactions on Smart Grid, 7(5), pp. 2397-2405, 2016.

[87] Shaker, Hamid, Hamidreza Zareipour, and David Wood, A data-driven approach for estimating the power generation of invisible solar sites, IEEE Transactions on Smart Grid, 7(5), pp. 2466-2476, 2016.

[88] Simmhan, Yogesh, Saima Aman, Alok Kumbhare, Rongyang Liu, Sam Stevens, Qunzhi Zhou, and Viktor Prasanna, Cloud-based software platform for big data analytics in smart grids, Computing in Science and Engineering, 15(4), pp. 38-47, 2013.

[89] Markovic, Dragan S., Dejan Zivkovic, Irina Branovic, Ranko Popovic, and Dragan Cvetkovic, Smart power grid and cloud computing, Renewable and Sustainable Energy Reviews, 24, pp. 566-577, 2013.

[90] Shyam, R., Bharathi Ganesh HB, Sachin Kumar, Prabaharan Poornachandran, and K. P. Soman, Apache spark a big data analytics platform for smart grid, Procedia Technology, 21, pp. 171-178, 2015.

[91] Hou, Weigang, Zhaolong Ning, Lei Guo, and Xu Zhang, Temporal, Functional and Spatial Big Data Computing Framework for Large-Scale Smart Grid, IEEE Transactions on Emerging Topics in Computing, 2017.

[92] Jang, Dongsik, Jiyong Eom, Min Jae Park, and Jae Jeung Rho, Variability of electricity load patterns and its effect on demand response: A critical peak pricing experiment on Korean commercial andindustrial customers, Energy Policy, 88, pp. 11-26. 2016.

[93] Shahzadeh, Abbas, Abbas Khosravi, and Saeid Nahavandi, Improving load forecast accuracy by clustering consumers using smart meter data, International Joint Conference on Neural Networks, pp. 1-7. IEEE, 2015.

[94] Liu, Xiufeng, Lukasz Golab, and Ihab F. Ilyas, SMAS: A smart meter data analytics system, 31st International Conference on Data Engineering, pp. 1476-1479. IEEE, 2015.

[95] Pan, Erte, Husheng Li, Lingyang Song, and Zhu Han, Kernel-based nonparametric clustering for load profiling of big smart meter data, In Wireless Communications and Networking Conference (WCNC), 2015 IEEE, pp. 2251-2255, IEEE, 2015.

[96] Panapakidis, Ioannis, Minas Alexiadis, and Grigoris Papagiannis, Evaluation of the performance of clustering algorithms for a high voltage industrial consumer, Engineering Applications of Artificial Intelligence, 38, pp. 1-13, 2015.

[97] Lavin, Alexander, and Diego Klabjan., Clustering time-series energy data from smart meters, Energy efficiency, 8(4), pp. 681-689, 2015.

[98] Green, Richard, Iain Staffell, and Nicholas Vasilakos, Divide and Conquer? k-Means Clustering of Demand Data Allows Rapid and Accurate Simulations of the British Electricity System, IEEE Transactions on Engineering Management, 61(2), pp. 251-260, 2014.

[99] Jarrah Nezhad, Aylin, Tri Kurniawan Wijaya, Matteo Vasirani, and Karl Aberer, SmartD: Smart meter data analytics dashboard, 5th international conference on Future energy systems, pp. 213-214. ACM, 2014.

[100] Miller, Clayton, Zoltn Nagy, and Arno Schlueter, A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings, Renewable and Sustainable Energy Reviews, 2017.

[101] Shao, Huijuan, Manish Marwah, and Naren Ramakrishnan, A temporal motif mining approach to unsupervised energy disaggregation, 1st International Workshop on Non-Intrusive Load Monitoring, Pittsburgh, PA, USA, 2012.

[102] Chicco, Gianfranco, Octavian-Marcel Ionel, and Radu Porumb, Electrical load pattern grouping based on centroid model with ant colony clustering, IEEE Transactions on Power Systems, 28(2), pp. 1706-1715, 2013.

[103] Borgeson, Samuel Dalton. Targeted efficiency: Using customer meter data to improve efficiency program outcomes. University of California, Berkeley, 2013.

[104] Chen, Zhonghui, Siying Chen, and Xinxin Feng, A design of distributed storage and processing system for Internet of Vehicles, 8th International Conference on Wireless Communications and Signal Processing, pp. 1-5. IEEE, 2016.

[105] Zheng, Xinhu, Wei Chen, Pu Wang, Dayong Shen, Songhang Chen, Xiao Wang, Qingpeng Zhang, and Liuqing Yang, Big data for social transportation, IEEE Transactions on Intelligent Transportation Systems, 17(3), pp. 620-630, 2016.

[106] Sankaranarayanan, Hari Bhaskar, and Ravish Singh Thind, Multi-modal travel in India: A big data approach for policy analytics, 7th International Conference on Cloud Computing, Data Science and Engineering-Confluence, pp. 243-248. IEEE, 2017.

[107] Hua, Xingcheng, Jierui Wang, Li Lei, Bin Zhou, Xiaolin Zhang, and Peng Liu, H-TDMS: A System for Traffic Big Data Management, pp. 85-96. Springer, 2016.

[108] Massobrio, Renzo, Andrs Pas, Nicols Vzquez, and Sergio Nesmachnow, Map-reduce for processing GPS data from public transport in Montevideo, Uruguay, In Simposio Argentino de GRANdes DAtos (AGRANDA 2016)-JAIIO, 45, 2016.

[109] Xiao, Zhe, Loganathan Ponnambalam, Xiuju Fu, and Wanbing Zhang, Maritime Traffic Probabilistic Forecasting Based on Vessels' Waterway Patterns and Motion Behaviors, IEEE Transactions on Intelligent Transportation Systems, 2017.

[110] Tselentis, Dimitrios I., George Yannis, and Eleni I. Vlahogianni, Innovative motor insurance schemes: a review of current practices and emerging challenges, Accident Analysis and Prevention, 98 pp. 139-148, 2017.

[111] Qin, Guoyang, Tienan Li, Bin Yu, Yunpeng Wang, Zhenhua Huang, and Jian Sun, Mining factors affecting taxi drivers incomes using GPS trajectories, Transportation Research Part C: Emerging Technologies, 79 pp. 103-118, 2017.

[112] Arbabzadeh, Nasim, and Mohsen Jafari, A Data-Driven Approach for Driving Safety Risk Prediction Using Driver Behavior and Roadway Infor-

mation Data, IEEE Transactions on Intelligent Transportation Systems, 2017.

[113] Schlgl, Matthias, and Rainer Sttz, Methodological considerations with data uncertainty in road safety analysis, Accident Analysis and Prevention, 2017.

[114] Rathore, M. Mazhar, Awais Ahmad, Anand Paul, and Gwanggil Jeon, Efficient graph-oriented smart transportation using internet of things generated big data, 11th International Conference on Signal-Image Technology and Internet-Based Systems, pp. 512-519. IEEE, 2015.

[115] Lv, Yisheng, Yanjie Duan, Wenwen Kang, Zhengxi Li, and Fei-Yue Wang, Traffic flow prediction with big data: a deep learning approach, IEEE Transactions on Intelligent Transportation Systems, 16(2) pp. 865-873. 2015.

[116] Fan, Cheng, Fu Xiao, and Chengchu Yan, A framework for knowledge discovery in massive building automation data and its application in building diagnostics, Automation in Construction, 50, pp. 81-90. 2015.

[117] Fan, Cheng, Fu Xiao, Henrik Madsen, and Dan Wang, Temporal knowledge discovery in big BAS data for building energy management, Energy and Buildings, 109, pp. 75-89, 2015

[118] Koseleva, Natalija, and Guoda Ropaite, Big Data in Building Energy Efficiency: Understanding of Big Data and Main Challenges, Procedia Engineering, 172, pp. 544-549, 2017.

[119] Fan, Cheng, and Fu Xiao, Mining big building operational data for improving building energy efficiency: A case study, Building Services Engineering Research and Technology, 2017.

[120] Li, Guannan, Yunpeng Hu, Huanxin Chen, Jiangyu Wang, Yabin Guo, Jiangyan Liu, and Jiong Li, Identification and isolation of outdoor fouling

faults using only built-in sensors in variable refrigerant flow system: A data mining approach, Energy and Buildings, 146, pp. 257-270, 2017.

[121] Bolchini, Cristiana, Angela Geronazzo, and Elisa Quintarelli, Smart buildings: A monitoring and data analysis methodological framework, Building and Environment, 121, pp. 93-105, 2017.

[122] Fagiani, M., Squartini, S., Severini, M. and Piazza, F, A novelty detection approach to identify the occurrence of leakage in smart gas and water grids. International Joint Conference on Neural Networks, pp. 1-8, IEEE, 2015.

[123] Reinhardt, Andreas, and Sebastian Koessler, PowerSAX: Fast motif matching in distributed power meter data using symbolic representations, IEEE 39th Conference on Local Computer Networks Workshops, pp. 531-538, IEEE, 2014.

[124] Iglesias, Flix, and Wolfgang Kastner, Analysis of similarity measures in times series clustering for the discovery of building energy patterns, Energies, 6(2), 579-597, 2013.

[125] Schlueter, Arno, Philipp Geyer, and Sasha Cisar, Analysis of georeferenced building data for the identification and evaluation of thermal microgrids, Proceedings of the IEEE, 104(4), pp. 713-725, 2016.

[126] Geyer, Philipp, Arno Schlter, and Sasha Cisar, Application of clustering for the development of retrofit strategies for large building stocks, Advanced Engineering Informatics, 31, pp. 32-47, 2017.

[127] Miller, Clayton, and Arno Schlueter, Forensically discovering simulation feedback knowledge from a campus energy information system, In Proceedings of the Symposium on Simulation for Architecture and Urban Design, Society for Computer Simulation International, pp. 136-143, 2015.

[128] Yarbrough, I., Q. Sun, D. C. Reeves, K. Hackman, R. Bennett, and D. S. Henshel, Visualizing building energy demand for building peak energy analysis, Energy and Buildings, 91, pp. 10-15, 2015.

[129] Miller, Clayton, Zoltn Nagy, and Arno Schlueter, Automated daily pattern filtering of measured building performance data, Automation in Construction, 49, pp. 1-17, 2015.

[130] Diong, B., G. Zheng, and M. Ginn, Establishing the foundation for energy management on university campuses via data analytics, SoutheastCon 2015, pp. 1-7. IEEE, 2015.

[131] Pieri, Stella Panayiota, and Mat Santamouris, Identifying energy consumption patterns in the Attica hotel sector using cluster analysis techniques with the aim of reducing hotels CO 2 footprint, Energy and Buildings, 94, pp. 252-262, 2015.

[132] Ploennigs, Joern, Bei Chen, Paulito Palmes, and Raymond Lloyd, e2-Diagnoser: A system for monitoring, forecasting and diagnosing energy usage, IEEE International Conference on Data Mining Workshop, pp. 1231-1234. IEEE, 2014.

[133] Georgescu, Michael, and Igor Mezic, Site-Level Energy Monitoring and Analysis using Koopman Operator Methods, 2014 ASHRAE/IBPSA-USA Building Simulation Conference (SimBuild 2014). Atlanta, GA, USA, 2014.

[134] Heidarinejad, Mohammad, Matthew Dahlhausen, Sean McMahon, Chris Pyke, and Jelena Srebric, Cluster analysis of simulated energy use for LEED certified US office buildings, Energy and Buildings, 85, pp. 86-97, 2014.

[135] Sun, Yunchuan, Houbing Song, Antonio J. Jara, and Rongfang Bie, Internet of things and big data analytics for smart and connected communities, IEEE Access, 4, pp. 766-773, 2016.

[136] Risteska Stojkoska, Biljana, Kire Trivodaliev, and Danco Davcev, Internet of Things Framework for Home Care Systems, Wireless Communications and Mobile Computing, 2017.

[137] Kazarian, A., V. Teslyuk, I. Tsmots, and M. Mashevska, Units and structure of automated smart house control system using machine learning al-

gorithms, 14th International Conference on Experience of Designing and Application of CAD Systems in Microelectronics, pp. 364-366. IEEE, 2017.

[138] Apostol, Elena, Catalin Leordeanu, Mariana Mocanu, and Valentin Cristea, Towards a Hybrid Local-Cloud Framework for Smart Farms, 20th International Conference on Control Systems and Computer Science, pp. 820-824. IEEE, 2015.

[139] Leena, H. U., B. G. Premasudha, and P. K. Basavaraja, Sensible approach for soil fertility management using GIS cloud, International Conference on Advances in Computing, Communications and Informatics, pp. 2776-2781. IEEE, 2016.

[140] Yethiraj, N. G., and Noor Ayesha, A Study to Improve Crop Yield in Agriculture Using IOT and Big Data, Adarsh Journal of Information Technology, 6(1), pp. 6-10, 2017.

[141] Tan, Li, Ronald Haley, Riley Wortman, Yiannis Ampatzidis, and Matthew Whiting, An integrated cloud-based platform for labor monitoring and data analysis in precision agriculture, IEEE 14th International Conference on Information Reuse and Integration, pp. 349-356. IEEE, 2013.

[142] Janssen, Sander JC, Cheryl H. Porter, Andrew D. Moore, Ioannis N. Athanasiadis, Ian Foster, James W. Jones, and John M. Antle, Towards a new generation of agricultural system data, models and knowledge products: Information and communication technology, Agricultural Systems, 155, pp. 200-212, 2017.

[143] Khanal, Sami, John Fulton, and Scott Shearer, An overview of current and potential applications of thermal remote sensing in precision agriculture, Computers and Electronics in Agriculture, 139, pp. 22-32, 2017.

[144] Xie, Hualin, Yafen He, and Xue Xie., Exploring the factors influencing ecological land change for China's BeijingTianjinHebei region using big data, Journal of Cleaner Production, 142, pp. 677-687, 2017.

[145] Alves, Gabriel M., and Paulo E. Cruvinel, Big Data environment for agricultural soil analysis from CT digital images, IEEE Tenth International Conference on Semantic Computing, pp. 429-431. IEEE, 2016.

[146] Wolfert, Sjaak, Lan Ge, Cor Verdouw, and Marc-Jeroen Bogaardt, Big Data in Smart FarmingA review, Agricultural Systems, 153, 69-80, 2017.

[147] Joshi, Jetendra, Siddhanth Polepally, Pranith Kumar, Rohith Samineni, S. R. Rahul, Kaushal Sumedh, Dandu Geet Kamal Tej, and Vishal Rajapriya, Machine Learning Based Cloud Integrated Farming, International Conference on Machine Learning and Soft Computing, pp. 1-6. ACM, 2017.

[148] Hot, Elma, and Vesna Popovi-Bugarin, Soil data clustering by using K-means and fuzzy K-means algorithm, In Telecommunications Forum Telfor (TELFOR), pp. 890-893. IEEE, 2015.

[149] Divya, M., T. N. Manjunath, and Ravindra S. Hegadi, A Study on Developing Analytical Model for Groundnut Pest Management using Data Mining Techniques, International Conference on Computational Intelligence and Communication Networks, pp. 691-696. IEEE, 2014.

[150] Gandhi, Niketa, and Leisa J. Armstrong, Rice crop yield forecasting of tropical wet and dry climatic zone of India using data mining techniques, IEEE International Conference on Advances in Computer Applications, pp. 357-363. IEEE, 2016.

[151] Gandhi, Niketa, and Leisa J. Armstrong, A review of the application of data mining techniques for decision making in agriculture, 2nd International Conference on Contemporary Computing and Informatics, pp. 1-6. IEEE, 2016.

[152] Ahrary, Alireza, and R. Dennis A. Ludena, Big data application to the vegetable production and distribution system, IEEE 10th International Colloquium on Signal Processing and its Applications, pp. 20-24. IEEE, 2014.

[153] He, Chenguang, Xiaomao Fan, and Ye Li., Toward ubiquitous healthcare services with a novel efficient cloud platform, IEEE Transactions on Biomedical Engineering, 60(1), pp. 230-234, 2013.

[154] Valenza, Gaetano, Luca Citi, Claudio Gentili, Antonio Lanata, Enzo Pasquale Scilingo, and Riccardo Barbieri, Characterization of depressive states in bipolar patients using wearable textile technology and instantaneous heart rate variability assessment, IEEE Journal of biomedical and health informatics, 19(1), pp. 263-274, 2015

[155] Swangnetr, Manida, and David B. Kaber, Emotional state classification in patientrobot interaction using wavelet analysis and statistics-based feature selection, IEEE Transactions on Human-Machine Systems, 43(1), pp. 63-75, 2013.

[156] Ng, Joseph Kee-Yin, Jiantao Wang, Kam-Yiu Lam, Calvin Ho Chuen Kam, and Song Han, Capturing and analyzing pervasive data for SmartHealth, 28th International Conference on Advanced Information Networking and Applications, pp. 985-992. IEEE, 2014.

[157] Rao, A. Ravishankar, and Daniel Clarke, An open-source framework for the interactive exploration of Big Data: Applications in understanding health care, International Joint Conference on Neural Networks, pp. 1641-1648. IEEE, 2017.

[158] Yassine, Abdulsalam, Shailendra Singh, and Atif Alamri, Mining Human Activity Patterns from Smart Home Big Data for Healthcare Applications, IEEE Access, 2017.

[159] Wang, Hai, Zeshui Xu, and Witold Pedrycz, An overview on the roles of fuzzy set techniques in big data processing: Trends, challenges and opportunities, Knowledge-Based Systems, 118, pp. 15-30, 2017.

[160] Dai, Xiangfeng, Marwan Bikdash, and Bradley Meyer, From social media

to public health surveillance: Word embedding based clustering method for twitter classification, SoutheastCon, pp. 1-7. IEEE, 2017.

[161] Zhong, Hongye, and Jitian Xiao, Enhancing Health Risk Prediction with Deep Learning on Big Data and Revised Fusion Node Paradigm, Scientific Programming, 2017.

[162] Hillerman, Tiago, Joo Carlos F. Souza, Ana Carla B. Reis, and Rommel N. Carvalho, Applying clustering and AHP methods for evaluating suspect healthcare claims, Journal of Computational Science, 19, 97-111, 2017.

[163] Quintana, David, Alejandro Cervantes, Yago Saez, and Pedro Isasi, Clustering technique for large-scale home care crew scheduling problems, Applied Intelligence, pp. 1-13, 2017.

[164] Bruse, Jan L., Maria A. Zuluaga, Abbas Khushnood, Kristin Mcleod, Hopewell N. Ntsinjana, Tain-Yen Hsia, Maxime Sermesant, Xavier Pennec, Andrew M. Taylor, and Silvia Schievano, Detecting Clinically Meaningful Shape Clusters in Medical Image Data: Metrics Analysis for Hierarchical Clustering applied to Healthy and Pathological Aortic Arches, IEEE Transactions on Biomedical Engineering, 2017 .

[165] Amirian, Pouria, Francois van Loggerenberg, Trudie Lang, Arthur Thomas, Rosanna Peeling, Anahid Basiri, and Steven Goodman, Using big data analytics to extract disease surveillance information from point of care diagnostic machines, Pervasive and Mobile Computing, 2017.

[166] Pan, Gang, Guande Qi, Wangsheng Zhang, Shijian Li, Zhaohui Wu, and Laurence Tianruo Yang, Trace analysis and mining for smart cities: issues, methods, and applications, IEEE Communications Magazine 51(6), pp. 120-126, 2013.

[167] Stolfi, Daniel H., Enrique Alba, and Xin Yao, Predicting Car Park Occupancy Rates in Smart Cities, International Conference on Smart Cities, pp. 107-117. Springer, 2017.

[168] Khan, Zaheer, Ashiq Anjum, and Saad Liaquat Kiani, Cloud based big data analytics for smart future cities, IEEE/ACM 6th international conference on utility and cloud computing, pp. 381-386. IEEE Computer Society, 2013.

[169] Thomas, Robert W., and Jos M. Vidal, Toward detecting accidents with already available passive traffic information, IEEE 7th Annual Computing and Communication Workshop and Conference, pp. 1-4. IEEE, 2017.

[170] Zheng, Yanxu, Sutharshan Rajasegarar, Christopher Leckie, and Marimuthu Palaniswami, Smart car parking: temporal clustering and anomaly detection in urban car parking, IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing, pp. 1-6. IEEE, 2014.

[171] Piovesan, Nicola, Leo Turi, Enrico Toigo, Borja Martinez, and Michele Rossi, Data Analytics for Smart Parking Applications, Sensors, 16(10), 2016.

[172] Yavari, Ali, Prem Prakash Jayaraman, and Dimitrios Georgakopoulos, Contextualised service delivery in the Internet of Things: Parking recommender for smart cities, IEEE 3rd World Forum on Internet of Things, pp. 454-459. IEEE, 2016.

[173] Kumar, S. Shathish, and A. Jayachandran, Detecting and Alerting Damaged Roads Using Smart Street System, 2017.

[174] Enrquez, Fernando, Luis Miguel Soria, Juan Antonio lvarez-Garca, Francisco Velasco, and Oscar Dniz, Existing Approaches to Smart Parking: An Overview, International Conference on Smart Cities, pp. 63-74. Springer, 2017.

[175] Xie, Xuemei, Chenye Wang, Shu Chen, Guangming Shi, and Zhifu Zhao, Real-Time Illegal Parking Detection System Based on Deep Learning, In-

ternational Conference on Deep Learning Technologies, pp. 23-27. ACM, 2017.

[176] Liu, Zhanyu, Research on the Internet of Things and the development of smart city industry based on big data, Cluster Computing, pp. 1-7, 2017.

[177] Ang, Li-Minn, Kah Phooi Seng, Adamu Zungeru, and Gerald Ijemaru, Big Sensor Data Systems for Smart Cities, IEEE Internet of Things Journal, 2017.

[178] Montori, Federico, Luca Bedogni, and Luciano Bononi, A Collaborative Internet of Things Architecture for Smart Cities and Environmental Monitoring, IEEE Internet of Things Journal, 2017.

[179] Alam, Furqan, Rashid Mehmood, Iyad Katib, Nasser Albogami, and Aiiad Albeshri, Data Fusion and IoT for Smart Ubiquitous Environments: A Survey, IEEE Access, 2017.

[180] Trilles, Sergio, scar Belmonte, Sven Schade, and Joaqun Huerta, A domain-independent methodology to analyze IoT data streams in real-time. A proof of concept implementation for anomaly detection from environmental data, International Journal of Digital Earth, 10(1), pp. 103-120, 2017.

[181] Shao, Zhenfeng, Jiajun Cai, and Zhongyuan Wang, Smart Monitoring Cameras Driven Intelligent Processing to Big Surveillance Video Data, IEEE Transactions on Big Data, 2017.

[182] Hashem, Ibrahim Abaker Targio, Ibrar Yaqoob, Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani, and Samee Ullah Khan, The rise of big data on cloud computing: Review and open research issues, Information Systems, 47, pp. 98-115, 2015.

[183] Kremer, Jan, Kristoffer Stensbo-Smidt, Fabian Gieseke, Kim Steenstrup Pedersen, and Christian Igel, Big Universe, Big Data: Machine Learning and Image Analysis for Astronomy, IEEE Intelligent Systems, 32(2), pp. 16-22, 2017

[184] Vakali, Athena, Leonidas Anthopoulos, and Srdjan Krco, Smart Cities Data Streams Integration: experimenting with Internet of Things and social data flows, 4th International Conference on Web Intelligence, Mining and Semantics, ACM, 2014.

[185] Chi, Qingping, Hairong Yan, Chuan Zhang, Zhibo Pang, and Li Da Xu, A reconfigurable smart sensor interface for industrial WSN in IoT environment, IEEE Transactions on Industrial Informatics, 10(2) pp. 1417-1425, 2014.

[186] Tao, Fei, Ying Cheng, Li Da Xu, Lin Zhang, and Bo Hu Li, CCIoT-CMfg: cloud computing and internet of things-based cloud manufacturing service system, IEEE Transactions on Industrial Informatics, 10(2), pp. 1435-1442, 2014.

[187] Jiang, Lihong, Li Da Xu, Hongming Cai, Zuhai Jiang, Fenglin Bu, and Boyi Xu, An IoT-oriented data storage framework in cloud computing platform, IEEE Transactions on Industrial Informatics 10(2), pp. 1443-1451, 2014.

[188] Lee, Jay, Hung-An Kao, and Shanhu Yang, Service innovation and smart analytics for industry 4.0 and big data environment, Procedia Cirp 16, pp. 3-8, 2014.

[189] Ryan, Peter J., and Richard B. Watson, Research Challenges for the Internet of Things: What Role Can OR Play?, Systems, 5(1), 2017.

[190] Ray, Partha Pratim, Towards an internet of things based architectural framework for defence, International Conference on Control, Instrumentation, Communication and Computational Technologies, pp. 411-416. IEEE, 2015.

[191] Puthal, Deepak, Surya Nepal, Rajiv Ranjan, and Jinjun Chen, Threats to networking cloud and edge data centers in the Internet of Things, IEEE Cloud Computing, 3(3), pp. 64-71, 2016.

[192] Kumar, Gunupudi Rajesh, Nimmala Mangathayaru, and Gugulothu Narsimha,A feature clustering based dimensionality reduction for intrusion detection, IADIS International Journal on Computer Science and Information Systems, 12(1), 2017.

[193] Raza, Shahid, Linus Wallgren, and Thiemo Voigt, SVELTE: Real-time intrusion detection in the Internet of Things, Ad hoc networks, 11(8), pp. 2661-2674, 2013.

[194] Hashem, Ibrahim Abaker Targio, Ibrar Yaqoob, Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani, and Samee Ullah Khan, The rise of big data on cloud computing: Review and open research issues, Information Systems, 47, pp.98-115, 2015.

[195] MacDougall, Daniel, and Warren B. Crummett, Guidelines for data acquisition and data quality evaluation in environmental chemistry, Analytical Chemistry, 52(14) pp. 2242-2249, 1980.

[196] Saha, Barna, and Divesh Srivastava, Data quality: The other face of big data, 30th International Conference on Data Engineering, pp. 1294-1297. IEEE, 2014.

Appendix.

Table .4: Analysis and classification of Big Data approaches applied in Healthcare, Energy, Transportation, Building automation, Smart Cities, Agriculture, Industry and Military IoT

Big Data				
	Storage	Cleaning/Cleansing	Analysis/Analytics	Visualization
Healthcare (35 papers)	Raspberry Pi [11], Cloud [13, 14, 15, 154, 153], HBase[6, 161], Spark SQL[6], Hive [6], Cloudera Impala [6], Apache Parquest [6], Python-based [157], Mysql [158], MongoDB Databases [158], NoSQL[159, 161], Centralized Database [165]	Integrity [15, 154], Information Exploration [161], Anomaly Detection [157, 158, 162, 164], Data Integration [159, 161], Reasoning [159], Co-reference [159], Resolution [159], Entity Linking [159], Information Extraction [159, 165, 161], Consolidation [159], Paraphrase [159], Resolution [159], Ontology Alignment [159], Information Fusion [159, 161], Logical Inference [161], Data Transformation [161], Aggregation [161, 159, 163, 160]	Semantic Analysis [11], Clustering [161, 157, 159, 158, 162, 6, 164, 165, 163, 160, 13], AHP Method [162], Classification [161, 157, 159, 162, 160, 13, 14] Feature Extraction [154, 14, 155, 156], MapReduce [161, 6], Strom [6], Hadoop [161, 6], Apache Mahout [6], Spark [6], Association rule [158], Patterns [161, 158], Fuzzy [159], Regression [162], Descriptive Statistics [162], Deep Learning [161], Decision Trees [161] , Bayesian Fusion [161], Neural Network [161], Apache Pig [6], Apache Flume [6], Apache Hadoop Yarn [6], Semantic Mining [159], Knowledge Discovery [165]	Data Visualization [161, 157, 159, 6][162, 164, 165, 13] Interpretation [161, 157, 159] Data Lineage [15] [154] Visual Exploration [157], Visual Analytics [165]

	Big Data			
	Storage	Cleaning/Cleansing	Analysis/Analytics	Visualization
Energy (24 papers)	Cloud [81, 82, 83, 85, 89, 78, 79, 88, 75, 23], Relational Databases [82, 86], NoSQL [81, 85, 86, 90], PostgreSQL[94], Microsoft SQLserver[81], XML Databases [81], Distributed File System [85]	Data Integration [81, 76, 85], Data Selection [84, 76, 77], Data Conversion [76], Outlier Detection [81, 82, 76, 77], Data Transformation [81, 77], Data Reduction [84, 87, 81], Data Discretization[81], Bad Data Detection [82, 86, 91], Separation [82], Motif Detection [101], Data Ingest Pipeline [88], Data validation and calibration [82], Outlier Data Identification [76][77], Disaggregation [101],	MapReduce [82, 75, 23], Hadoop [84, 85], Spark [84, 85, 86, 90], Storm [85], Flink [85], Classification [81, 84, 90, 75], Clustering [81, 82, 84, 85, 86, 87, 88, 90, 91, 92, 93, 95, 96, 97, 98, 102, 103, 76, 77, 80] Motif Mining [101], Dimensionality Reduction [75], Neural Network [82, 76], Fuzzy [82, 76], Gap Statistic Algorithm [76], Frequent Pattern Mining [81], Association Rule Mining [81, 90], Pre-clustering [84], Analytic Hierarchy Process [85], Regression[81, 90], Forecasting[81, 82, 88]	Data Visualization [82, 84, 85, 86, 88] Visual Analytics [94, 99]
Transportation (11 papers)	Cloud [107, 110, 25], NoSQL [104], Hbase [107, 104],	Data Collection [105, 107, 108, 109, 110, 111, 113, 114], Bad Data detection [105, 107, 108, 109, 111, 112], Data Ingestion [106], Data Integration [107], Data Preparation [113], Data Observation [113], Data Fusion [105], data Aggregation [109, 112, 113, 114, 26]	Neural Network [109, 26, 115], Deep Learning [107, 26, 115], Clustering [105, 106, 107, 109, 113, 25], Fuzzy [106], Genetic Algorithm [28], Streaming Analysis [29], Hadoop [104, 107, 114], Spark [104, 107], classification [105, 106, 107], MapReduce [107, 108], Regression [105, 106, 109, 110, 111, 112, 113]	Data Visualization [105, 106, 107, 111, 112, 113, 114, 26], Visual analytics [105]

Big Data				
	Storage	Cleaning/Cleansing	Analysis/Analytics	Visualization
Building automation (20 papers)	Cloud [100, 135, 136], NoSQL [135], Excel [130], Model Database DB2RDF [132],	Hampel Filter [117], Data Exploration [116], Novelty Detection [100, 122, 132], Data Identification and Gathering [130], Tagging and Filtering [129], Discord Detection [100], Bad Data Detection [121], Data observation [120], Data Quality Screening [127], Sensitivity Screening [127], Breakout Detection Screening [127], Data Preparation [126], Data Collection [121, 100], Redundancy [121], Sampling [121], Data Modeling [132], Aggregation [116, 117, 121, 123, 125, 126, 127, 128, 129], Disaggregation [100, 122, 133]	Neural Network [137], Clustering [116, 117, 119, 120, 121, 100, 122, 123, 124, 125, 126, 127, 129, 131, 134], Segmentation [121], Fuzzy [125], Classification [120, 121, 100, 131], Decision Tree [119, 120], Multiple Linear Regression [120], Post-mining [116, 117], Association Rule Mining [117, 119, 121], Knowledge Discovery [116, 117, 119], Rule Extraction [100], Regression [121, 100, 133], Pattern Mining [132] Social Media Analytics [118]	Data Visualization [116, 121, 100, 123, 126, 127, 128, 129, 130, 131, 132, 133, 134], Interpretation [129, 132], Visual Analytics [100, 128, 130, 133]

Big Data				
	Storage	Cleaning/Cleansing	Analysis/Analytics	Visualization
Smart Cities (18 papers)	Cloud [172, 176, 177, 178, 180, 181], NoSQL [168, 177, 40], Apache Cassandra [168, 177], Hbase [173, 177], CouchDB [177], MongoDB [177], Big Table [177],	ETL [40], ontology [168, 41], Data Detection [166, 177], Data Transformation [167], Data Parser [167], Data Integration [168], Data Extraction [168, 177], Anomaly Detection [170, 171, 173, 174, 175, 179, 180, 181], Data Observation [172, 177, 179, 184], Data Fusion [173, 179] Summarization [184] Aggregation [166, 177, 178, 179, 180] Contextual Aggregation [172]	ThingSpeak [178], Spark [178], Neural Network [173, 179, 41, 177], Classification [166, 168, 169, 171, 173, 174, 175, 178, 179], Ranking [166], Regression [166, 169, 173, 177], Clustering [166, 167, 168, 170, 41, 171, 173, 174, 175, 176, 177, 178, 180, 181, 179], Kernel Methods [170], Data Inference [177], Support Vector Machines [173, 179], Patterns [166, 173, 179] RapidMiner [168], Hadoop [168, 173, 177], Spark [177], MapReduce [173, 177], Naive Bayes [173, 179, 181], Decision Tree [173], K-nearest Neighbor [173], Fuzzy [174, 179, 181], Deep Learning [175, 179], Conditional Random [177], Maximization [177], Contextual Filter [172], Knowledge Discovery [173], Semantic Networks [184]	Data Visualization [166, 167, 168, 169, 171, 173, 174, 175, 179, 180, 181, 184, 40]

Big Data				
	Storage	Cleaning/Cleansing	Analysis/Analytics	Visualization
Agriculture (13 papers)	Cloud [141, 138, 139], PostgreSQL [149]	Data Fusion [147, 140], Data Exploration [142, 149], Semantic Ontologies [142], Bad Data Detection [143], Data Integration [146, 152, 142], Anomaly Detection [148], Data Observation [149], Data Reduction and Rejection [149], Data Selection [149], Completeness Checking [142], Semantic Alignment [142], Aggregation[142, 144], Verification [138]	Classification [147, 148, 150, 151, 140], Filtering [138], Regression [143, 144, 147, 148, 149, 150, 151], Hadoop [145], Bayesian Network[147, 151], Association Rule Mining [148, 149, 151], Sequence [148], Clustering [148, 149, 150, 151], Fuzzy [148], Neural Network [151]	Data Visualization [141, 142, 143, 149, 151, 152] Interpretation [142, 145, 146, 147, 148, 151], Visual Analytics [142]
Industry (10 papers)	Data Access [14], Cloud [14], NoSQL [187], Relational Database [187]	Data Transmission [51, 185, 186], Data Integration [51, 61, 186, 187], Data Confidentiality [61], Data Aggregation [188], Data Integrity [14], Data Fusion [186], Data Conversion [186]	Data Inference [61], knowledge Base [188], Bayesian Belief Network [188], semantic Data Model [14], Hadoop [187]	Data Presentation [51, 61], Interpretation [61]
Military (8 papers)	Cloud [189, 190, 191],	Data Integration [57, 58], Data Fusion [189, 192, 57], Data Validation [57], Data Quality [189, 192, 193, 57], Data Exchange [59, 57]	Data Filtering [57, 58], Clustering [57], Prioritization [58], Data Envelopment [189]	Data Visualization [190], Interpretation [190]

Dr. Mouzhi Ge is an Assistant Professor at the Faculty of Informatics of the Masaryk University in Czech Republic. Prior to that, he has worked as an Assistant Professor at the Free University of Bolzano in Italy, as a Senior Researcher at the Bundeswehr University Munich in Germany, as an IT Manager at the Oxbridge Investment Ltd. in the UK, as a Postdoctoral Researcher at the Technical University Dortmund in Germany, and as a Visiting Researcher at the University of Milano Bicocca in Italy. He obtained his Ph.D. from the Dublin City University in Ireland in 2009. His research is mainly focused on Big Data, Intelligent Systems, IoT, and Data Management. He has been served as a program committee member and reviewer for different conferences (such as WWW, ICIS, ECIS, SIGIR, ACM SAC, ACM UMAP, EDBT, RecSys etc.) and journals (such as TOIS, IEEE TKDE and JECR etc.). His papers have been published in various International Journals such as *International Journal of Human-Computer Studies*, *Journal of Computer Information Systems*, and *International Journal of Semantic Computing*.

Hind Bangui is currently a Ph.D student with the Applied Mathematics and Computer Science Laboratory of the Cadi Ayyad University (UCAM) in Morocco. Her research is mainly focused on Mobile cloud computing. She continues with related topic as an exchange Erasmus Ph.D student with the Laboratory of Software Architectures and Information Systems (LaSArIS) at the Faculty of Informatics of the Masaryk University in Czech Republic. At the same time, she is involved in Big Data project with the Institute of Computer Science of the Masaryk University in Czech Republic.

Dr. Barbora Bühnová is an Assistant Professor in the Faculty of Informatics of the Masaryk University in Czech Republic. She received her Ph.D. degree in computer science at Masaryk University in Brno, Czech Republic in 2008, for application of formal methods in software architecture design. She continued with related topics as a postdoc researcher at University of Karlsruhe and Research Center for Information Technology (FZI) in Karlsruhe, Germany, and later at Swinburne University of Technology in Melbourne, Australia. She has been involved in organization of numerous top conferences (e.g. ICSE, ICSA, ESEC-FSE, QoSA, CBSE, MOBILESoft) and acts as a reviewer and guest editor in multiple journals (e.g. IEEE TSE, Elsevier SCP, Elsevier JSS, Springer SoSyM, Wiley SME). Her papers have been published in various International Journals such as *IEEE Transactions on Software Engineering*, *Journal of Systems and Software,* and *Software and Systems Modeling*.

Mouzhi Ge

Hind Bangui

Barbora Bühnová

- Analyzed the state-of-the-art Big Data research in IoT.
- Compared the Big Data technologies across eight IoT domains such as healthcare, transportation, agriculture etc,.
- Suggested which Big Data technologies can possibly be used in which IoT domain.
- Interpreted the Big Data research and application opportunities in eight IoT domains.
- Discussed the pros and cons of Big Data technologies in IoT domains.