

Review Article

Review of the Complexity of Managing Big Data of the Internet of Things

David Gil ¹, Magnus Johnsson ^{2,3,4}, Higinio Mora ¹, and Julian Szymański ⁵

¹University of Alicante, Alicante, Spain

²Malmö University, Malmö, Sweden

³Department of Intelligent Cybernetic Systems, NRNU MEPhI, Moscow, Russia

⁴AI Research AB, Höör, Sweden

⁵Gdansk University of Technology, Gdansk, Poland

Correspondence should be addressed to David Gil; david.gil@ua.es

Received 9 November 2018; Revised 4 January 2019; Accepted 15 January 2019; Published 3 February 2019

Academic Editor: Danilo Comminiello

Copyright © 2019 David Gil et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

There is a growing awareness that the complexity of managing Big Data is one of the main challenges in the developing field of the Internet of Things (IoT). Complexity arises from several aspects of the Big Data life cycle, such as gathering data, storing them onto cloud servers, cleaning and integrating the data, a process involving the last advances in ontologies, such as Extensible Markup Language (XML) and Resource Description Framework (RDF), and the application of machine learning methods to carry out classifications, predictions, and visualizations. In this review, the state of the art of all the aforementioned aspects of Big Data in the context of the Internet of Things is exposed. The most novel technologies in machine learning, deep learning, and data mining on Big Data are discussed as well. Finally, we also point the reader to the state-of-the-art literature for further in-depth studies, and we present the major trends for the future.

1. Introduction

The fast-developing and expanding area known as the Internet of Things (IoT) [1–3] involves expanding the Internet beyond such standard devices as computers, smartphones, and tablets to also include the connection of other physical devices and objects. This allows for a variety of devices, sensors, etc. to be monitored and controlled, and to interact and communicate via the Internet. This means that an abundance of opportunity for brand new and revolutionary types of services and applications arises. As a result, we are now witnessing a technological revolution where millions of people are connecting and generate tremendous amounts of data through the increasing use of a wide variety of devices. These include smart devices and any type of wearable that are connected to the Internet, powering novel connected applications and solutions. The cost of technology has sharply decreased making it possible for everybody to access the Internet and to gather data and an abundance of real-time information.

One immediate consequence of this revolutionary emergence of novel technological opportunities is the urgent need for the development and adaptation of other related areas to further enable the development of the IoT field. Thus, new words, as well as new expressions, have started to emerge, such as Big Data [4, 5], cloud computing [6], and Data science. Data science has been defined as a “concept to unify statistics, data analysis, machine learning and their related methods” to “understand and analyse actual phenomena” with data [7, 8], and there is now a strong demand for professional data scientists in a multitude of sectors [9–12].

This article aims at providing a review of IoT related surveys in order to highlight the opportunities and the challenges, as well as the state-of-the-art technologies related to Big Data. There will be a particular focus on how to address the arising problems of managing the ensuing increased complexity. Since it is such a complex area, we have divided the Big Data procedure into several different stages to establish the most important points in each, while highlighting to

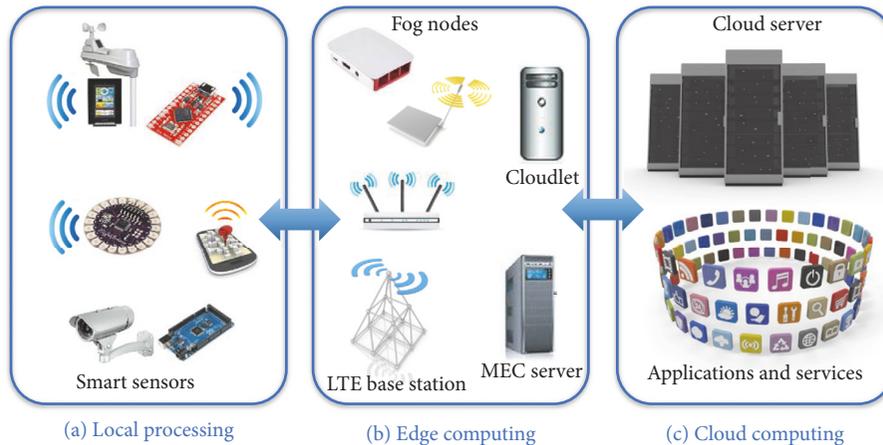


FIGURE 1: A schematic depiction of three different approaches to handle the complexity of the intensive data processing arising as a consequence of the tremendous amounts of collected IoT data.

the reader the most relevant papers related to every stage. Due to the complexity of managing Big Data, we have created separate sections in regard to the aforementioned stages of Big Data procedure. Our contribution explicitly indicates the advantages of every stage in the knowledge discovering procedure in contrast to approaches that offer more general visions. The advantage of this proposal is to be able to understand as well as analyse the challenges and opportunities in every particular phase.

The remainder of this article is structured as follows: first the next section discusses a set of general approaches to handle the complexity of managing Big Data in the context of the IoT as well as the future trends in the development of these approaches; then a section follows that discusses the knowledge discovering procedure in data gathered from a large number of diverse devices in the context of the IoT; finally, we provide a conclusion that summarises the article and points out major future trends.

2. The Internet of Things and Complexity Handling: Architectures for Big Data

The Internet of Things (IoT) paradigm has brought a great revolution to our society [13–15]. It is a technology that makes our world better. It allows us to get information about the physical environment around us, and from this data valuable knowledge can be inferred about how the world works. This knowledge enables the deployment of new real-world applications, and it makes it easier for smart decisions to improve the quality of life of the citizens of our society. There are many examples of how this novel technology runs. The smart city concept is a representative use case, where many applications have been developed for its ecosystem [16–19].

An important source of complexity within the IoT paradigm comes from the great amount of data collected. In most cases, the data also need to be processed in order to be converted into useful knowledge.

In view of the recent proposals on how to handle the complexity of Big Data, there are three general approaches

to carry out the ensuing very intensive data processing: (A) local processing; (B) edge computing; and (C) cloud computing. Figure 1 shows a schematic overview of these approaches, and Table 1 summarises a representative set of ways and aspects of handling the complexity arising from the IoT. Table 1 also provides references to corresponding papers, categorised under the headings of the three general approaches mentioned above. In the following subsections, brief descriptions of each of these approaches are presented, and finally their main future trends are introduced.

2.1. Local Processing. This approach basically consists of processing the data where the data is collected. In this way, no raw data need to be communicated to remote servers. Instead, only the useful and relevant information is centralised to make smart decisions [20, 21]. In addition, deploying the first-level intelligence closer to the sensors produces an increase in the overall energy efficiency and significantly reduces the communication needs of many IoT applications.

This approach develops the concept of ‘smart sensor,’ which was initially defined as ‘smart transducer’ [22]. A smart sensor is a sensor with computing and communication capabilities to make computations with the acquired data, make decisions, and store information for further use and perform two-way communications [23]. Smart sensors are becoming integral parts of intelligent systems and they are indispensable enablers of the IoT paradigm and the corresponding development of advanced applications. A typical example of these developed sensors is the ‘smart wearable.’ This device can acquire several biosignals, process them, show elaborated information to the user, and send the relevant information to, for example, external platforms for medical supervision [23–25]. Other important applications come from the logistics [26] and industrial fields [27]. Indeed, the new computation and communication capabilities of the IoT paradigm allow for the implementation of intelligent manufacturing systems giving rise to the next generation of industry, the so-called ‘Industry 4.0’ [28].

TABLE 1: A representative set of ways and aspects of handling the complexity arising from the IoT, together with references to corresponding papers, categorized under the headings of three general approaches (local processing, edge computing and cloud computing) to carry out intensive data processing of Big Data.

Work	Main contributions
(A) Local processing	
Smart sensing for IoT applications [21]	Discusses emerging trends of smart sensing.
Sensor Fusion and Smart Sensor in Sports and Biomedical Applications [24]	An overview of smart sensors and sensor fusion.
High-level modelling and synthesis of smart sensor networks for Industrial Internet of Things [27]	Efficient design process and methodology for complex industrial applications.
Smart Sensing Devices for Logistics Application [26]	Analysis of the logistics sector and Cyber Physical Systems (CPS) as smart connected solutions.
Intelligent Manufacturing in the Context of Industry 4.0: A Review [28]	Review of key technologies such as the IoT and cyber-physical systems.
(B) Edge computing	
Edge Computing: Vision and Challenges [34]	Challenges and opportunities in the field of edge computing are described.
Collaborative Working Architecture for IoT-Based Applications [20]	Network design, which combines sensing and processing capabilities based on the MCC paradigm.
IoT-Based Computational Framework [25]	Distributed framework based on the IoT paradigm for real-time monitoring.
Edge Computing [35]	Analysis of the edge computing paradigm.
Fog Computing [48]	The Fog Computing framework.
Secure Multi-Tier Mobile Edge Computing Model [49]	Formal framework to handle the security level of edge computing environments.
Mobile Edge Computing [38]	Analysis of opportunities, solutions, and challenges of the MEC paradigm.
Cloudlets [39]	Introduction to the cloudlet concept for offloading computations.
Future Edge Cloud and Edge Computing for Internet of Things Applications [40]	Discussion of Edge Cloud and Edge Computing research efforts.
(C) Cloud computing	
A Smart Sensing Architecture for Domestic Monitoring [50]	Integrated sensor network deployment with advanced Cloud Computing Data Mining algorithms.
IoT-as-a-Service [43]	Strategy for evaluating the information quality in delivering IoT-as-a-Service.
The shift to Cloud Computing [41]	This paper analyses the impact of the shift to the Cloud-based model.
Accessibility analysis in smart cities [46]	Comprehensive system for monitoring urban accessibility in smart cities.
Big data analytics framework for smart cities [47]	Smart City Data Analytics Panel for Big Data analytics.
Sensing and Actuation as a Service Delivery Model [44]	A novel system model for Sensing and Actuation as a Service (SAaaS).
User Quality-Of-Experience and Service Provider Profit in 5G [51]	The Quality-of-Experience (QoE) and the Profit-aware Resource Allocation problems are analysed.
Orchestrated Platform for Cyber-Physical Systems [45]	Discussion on the scalability of the sensor data back-end and the predictive simulation architecture for CPS.

In these environments, network virtualization plays a significant role in providing flexibility and better manageability to Internet [29]. This is a way for reducing the complexity of the infrastructure since network resources can be managed as logical services, rather than physical resources. This feature enables us to implement smart scheduling methods for network usage and dataflows routing from IoT applications [30].

In order to properly carry out this resource management, network performance monitoring needs to be performed

in effective and efficient ways. However, it remains a challenge for network operators [31] since active monitoring techniques used to dynamically acquire it can introduce overheads in the network [32]. In general, existing methods are hard to use in practice and further research is needed in this area. Nevertheless, a promising idea to address this challenge consists in reducing the data measurement by implementing intelligent measurement schemes based on inference techniques from partial direct monitoring data [33].

2.2. Edge Computing. Edge computing is a novel paradigm which has spawned great interest recently. It consists of the deployment of storage and computing capabilities at the ‘Edge’ of the Internet. The ‘Edge’ of the Internet can be defined as the portion of the network between sensors or data sources and cloud data centres [34]. The edge computing paradigm aims at deploying computing, storage, and network resources in this portion. The physical proximity of the computing platforms to where the data acquisition happens makes it easier to achieve lower end-to-end latency, high bandwidth, and low jitter to services [35].

There are several ways to implement edge computing that have in turn led to different approaches, such as Fog Computing, Mobile Edge Computing (MEC), and Cloudlet Deployment. Fog Computing consists in using the network devices such as routers, switches, and gateways as Fog Nodes to provide storage and computing resources [36]. In addition, network virtualization has significantly contributed to developing this paradigm by considering the fog devices as virtual network nodes. This trend increases the deployment flexibility of Fog Computing services and their integration with mobile devices and ‘things’ [37]. MEC is a novel paradigm based on deploying cloud computing capabilities in the base stations of the telecom operators [38]. Finally, Cloudlet Deployment consists in the same concept as Cloud Computing, but without the Wide Area Network (WAN) inconveniences. The servers are installed within the local networks where the data sources are connected. These servers are known as cloudlets [39].

Applications for edge computing, such as in Virtual Reality and Gaming Applications [40], cannot tolerate high latency, or its unpredictability. This is something that remote cloud servers cannot deliver.

2.3. Cloud Computing. The Cloud Computing paradigm is one of the most disruptive technological innovations in the last few years. It makes available to anyone a flexible amount of computing resources under per-use payment methods, the so-called ‘as-a-service’ model. Currently, more and more software and hardware solutions are redesigned for this cloud paradigm [41].

The cloud computing model favours the development of large-sized data centres where the resources are optimised through virtualization and efficient management systems. This technology gives the IoT applications the possibility to work in different environments in a very agile way using the same infrastructure [42]. In such a way, combining the cloud computing paradigm with IoT forms a new type of distributed system able to provide IoT-as-a-Service (IoTaaS) [43]. This concept allows for the integration of powerful computing resources with different types of devices such as sensors, actuators, and other embedded devices to deliver advanced services and applications based on the gathered data. A particular instance of this idea is the Sensing and Actuation Cloud where the connected IoT devices are mainly sensors and actuators [44], or the Cloud Cyber Physical Systems (CPS) composed of sensors or sensor networks [45].

There are a great variety of successful examples of this trend in many areas, where the data are analysed in the cloud

through Big Data and data mining methods to infer valuable knowledge from them and deliver rich and smart services to the stakeholders. For example, the smart city concept, mentioned above, is in part made possible by a centralised cloud-based data analysis and service provision [41, 46, 47].

In addition, a combination of these options can be designed taking several aspects into account, such as power consumption, communication networks, and the availability of computing platforms. Dynamic solutions can easily adapt to the more favourable approach to better handle the complexity and meet the operation constraints.

2.4. Future Trends. Regarding the future trends of the developments of these three general approaches to intensive data processing of IoT related Big Data, there are developments at several fronts. The following is a summary of those most relevant.

When it comes to local processing, the efforts are directed towards the continuous improvement of smart sensor devices. We can distinguish several research lines here. One is the efforts to increase the performance of the devices while simultaneously reducing their power consumption. Another is the integration of multiple sensing modalities on the same chip. Still another is the efforts directed towards the improvement of the methods employed for the extraction of useful information from the raw data [21].

Edge computing has a promising future since it decentralises the computing power along the network and produces clear benefits when it comes to response time and reliability [34]. The research lines in this field aim at reaching a smooth engagement with the IoT ecosystem, mainly by reducing the management complexity of dispersed edge resources and developing mechanisms to maintain the security perimeter for the data and applications [49].

The cloud computing paradigm has triggered a strong growth of computing services around the world. For this reason, there is intensive ongoing research on expanding cloud services and solutions to new fields of application. These tasks seek to simplify business and make services easier for stakeholders. In this way, the new 5G protocol will facilitate access for services and applications in the cloud improving the Quality-of-Experience [51].

3. Knowledge Discovering Procedure

In Figure 2, a classical procedure of discovering knowledge from the data gathered from a large number of diverse devices is depicted. In this figure, we get an overview of all the stages involved in such a process. There are many challenges involved in these stages that will be described next.

3.1. IoT Data Gathering. The gathering of data for IoT architectures involves collection from different sources like social networks, the web, various devices, software applications, humans, and not the least various kinds of sensors. In addition to physical sensors, there are also virtual sensors that are created by the combination and fusion of data from different physical sensors in the cloud [52]. When it

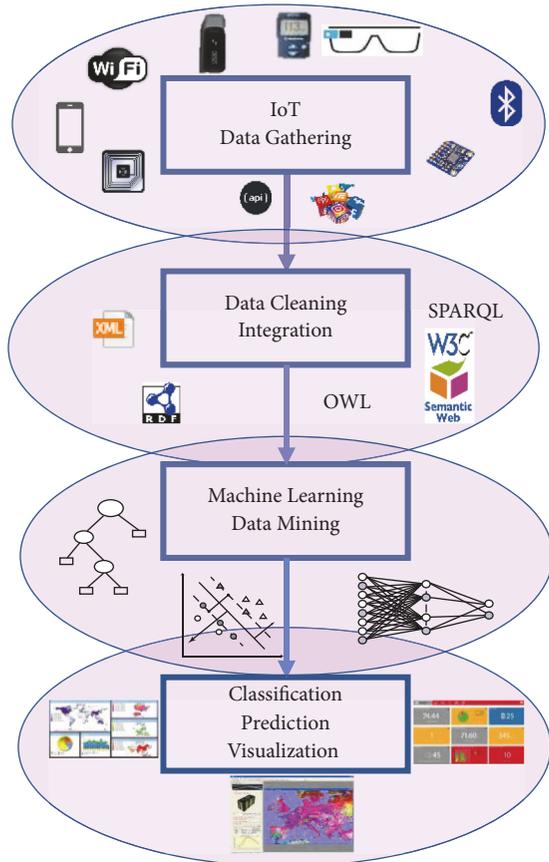


FIGURE 2: A classical procedure for the discovery of knowledge based on data gathered from a large number of diverse devices.

comes to the gathering of data from sensors, not only the raw sensor data are collected and stored, but these are also often linked to, for example, relevant contextual information, which increases the value of the data [53]. All these different sources engender large amounts of various types of data that, of course, also increases the requirements for storage capacity. The increasingly affordable storage resources that have recently become available mitigates this problem to some extent though.

Sensor networks are central for realising the IoT and in order to handle large amounts of polymorphous, heterogeneous sensor data on a large scale. Very Large-Scale Sensor Networks are employed using Cloud Computing [54]. Some of the main challenges regarding Very Large-Scale Sensor Networks are to handle the sensor resources and the computational resources and to store and process the sensor data.

Table 2 provides references to papers focused on the gathering of data in the context of the IoT.

3.2. Data Cleaning and Integration. A consequence of the way information is gathered through various sources and devices within IoT is that the information varies broadly in structure and type. This leads to a need for integration,

which can be defined as a set of techniques used to combine data from disparate sources into meaningful and valuable information.

Integration is one of the most challenging issues of Big Data, which is also associated with one of the most difficult Vs of Big Data, i.e., the variety of data. Table 3 shows a summary of papers that are focused on the problem of variety of information in Big Data.

Moreover, given the current context in which companies are organized, it is not enough to work with internal, local, and private databases. In most cases, there is also a need for the World Wide Web where many diverse databases and other data sources must interact and interoperate. This circumstance leads us to concepts such as heterogeneity and uncertainty.

Table 4 summarizes papers that deal with integration by means of a diversity of techniques and methods like XML, ontological constructs from knowledge representation, uncertainty, and data provenance.

3.3. Data Mining and Machine Learning. As more devices, sensors, etc. generate large amounts of data within the IoT, the question arises whether there are possibilities of finding hidden information in that data.

Data mining is a process that detects interesting knowledge from information repositories. This process is partly based on methods derived from modern machine learning algorithms adapted to fit Big Data and that extracts hidden information from, e.g., databases, data warehouses, data streams, time series, sequences, text, the web, and the large amount or valuable data generated by the IoT. Data mining aims at creating efficient predictive and descriptive models of large amounts of data that also generalize to new data [78]. It includes methods such as clustering, classification, time series analysis, association rule mining, and outlier analysis [79]. The precise choice among diverse data mining and machine learning techniques often depends on the taxonomy of the dataset.

Clustering includes unsupervised learning and uses the available structure to group data based on various kinds of similarity measures. Some examples of clustering methods are hierarchical clustering and partitioning algorithms, e.g., K-Means.

Classification is the process of finding models/functions describing classes that allow the prediction of class membership for new data. Some examples of classification methods are the K-Nearest Neighbour algorithm, Artificial Neural Networks, Decision Trees, Support Vector Machines, Bayesian Methods, and Rule-Based Methods.

In time series analysis meaningful properties are extracted from data over time, and in association rule mining, association rules are detected based on attribute-value conditions that are found frequently in the dataset.

Outlier analysis detects patterns that differ significantly from the main part of the data. The methods used are based on properties such as the density distribution or the distances between the instances in the data.

TABLE 2: The table summarizes and refers to a representative set of papers focusing on the gathering of data in the context of the IoT.

Reference	Title of paper	Description/Objective
[2]	Sensing as a Service and Big Data	Examines new approaches of IoT architectures, big sensor network applications, sensor data, and context-aware capturing techniques.
[55]	Internet of Things (IoT): A vision, architectural elements, and future directions	Describes an approach based on the cloud for worldwide implementation of Internet of Things.
[56]	Health monitoring and management using Internet-of-Things (IoT) sensing with cloud-based processing: Opportunities and challenges	Emphasis on the opportunities and challenges for the IoT and its future perspective in the health care area.
[15]	Internet of Things: A review of Surveys based on Context Aware Intelligent Services	A review of IoT studies that offer integrated and context-aware intelligent services.
[57]	Compressed sensing signal and data acquisition in wireless sensor networks and Internet of Things	Discusses how new insights can be supplied by compressed sensing into data sampling and acquisition for IoT.
[58]	A Computational Architecture Based on RFID Sensors for Traceability in Smart Cities	A novel approach of a distributed system to represent as well as providing the pathway and movement of people in densely geographical areas by means of a smart sensor network based on RFID.
[59]	Introducing a Novel Hybrid Artificial Intelligence Algorithm to Optimize Network of Industrial Applications in Modern Manufacturing	General modelling to evaluate and optimize nonlinear RFID network planning problems utilizing artificial intelligence techniques.

TABLE 3: The table summarizes and refers to a representative set of papers focusing on the variety of information in the context of the IoT.

Reference	Title of paper	Description/Objective
[60]	Data-intensive applications, challenges, techniques and technologies: A survey on Big Data	Survey on Big Data: applications, opportunities challenges, techniques and technologies.
[61]	The rise of “big data” on cloud computing: Review and open research issues	Important concepts of Big Data are introduced in this study and relationships among those concepts are provided. Finally, it summarizes the open research areas and how they need to be addressed.
[62]	Deep learning applications and challenges in big data analytics	In this paper, it is indicated how beneficial Deep Learning could be for several aspects of Big Data pattern recognition, analytics, semantic, etc.
[63]	On the use of MapReduce for imbalanced big data using Random Forest	In this experimental study, the performances with Random Forest classifier and MapReduce scheme have been used in order to deal with Imbalanced dataset.

Table 5 provides a summary of, and references to, papers focusing on machine learning and data mining in the context of Big Data.

3.4. Deep Learning. In recent years, deep learning has become an important technology for solving a wide range of machine learning tasks [85]. There are applications for natural language processing [86], signal processing [87], and video analysis that allows for the achievement of significantly better results than the state-of-the-art baselines. Also, deep learning is a very useful tool for processing large volumes of data [62]. Because of high efficiency of processing data obtained from complex sensing environments at different spatial and temporal resolutions, deep learning is a suitable tool for analysing real-world IoT data. According to Gartner’s Top 10 Strategic Technology Trends for 2017 ([https://www.gartner.com/smarterwithgartner/gartners-top-](https://www.gartner.com/smarterwithgartner/gartners-top-10-technology-trends-2017/)

[10-technology-trends-2017/](https://www.gartner.com/smarterwithgartner/gartners-top-10-technology-trends-2017/)), deep learning and IoT will become one of the most strategic technological two-way relationships: from the IoT side there are large volumes of data produced that require advanced analytics offered by the deep learning side. A wide range of deep learning architectures [88] finds applications for processing the data from IoT environments: convolutional networks for image analysis, recurrent networks for signal processing, autoencoders for denoising, feed forward networks for classification, and regression. Figure 3 represents a general architecture of deep learning.

Usually, the data are processed in dedicated frameworks such as Tensorflow (<https://www.tensorflow.org/>), Theano (<http://deeplearning.net/software/theano/>), Caffe (<http://caffe.berkeleyvision.org/>), H2O (<https://www.h2o.ai/>), and Torch (<http://torch.ch/>). Often GPUs or clusters of GPU servers are used for the processing [78, 79].

TABLE 4: The table summarizes and refers to a representative set of papers focusing on Data Integration in the context of the IoT.

Reference	Title of paper	Description/Objective
[64]	Principles of Data Integration	This paper brings up the notion that new services have to be able to share data among several applications and organizations, as well as integrating the data efficiently and flexibly.
[65]	Answering queries using views: A survey	This paper presents a survey of important methods that are employed to answer queries using views.
[66]	MiniCon: A scalable algorithm for answering queries using views	In this paper a survey of methods for efficient and comprehensive answering of queries using views is presented.
[67]	XQuery: the XML query language	This paper introduces the XML query language XQuery.
[68]	From semistructured data to XML: Migrating the Lore data model and query language	This paper discusses the adaptation to XML of databases and semistructured languages.
[69]	Querying XML streams	In this paper a construct called TurboXPath, similar to x-scan, is used for processing hierarchical “native XML” data pages written to disk.
[70]	Semantic integration: a survey of ontology-based approaches	This paper provides a survey of ontology-based approaches to semantic integration.
[71]	Learning to map between ontologies on the semantic web	This paper presents assisting tools for the mapping between ontologies on the semantic web.
[72]	Containment of conjunctive queries on annotated relations	This paper indicates the relationships between different provenance formalisms.
[73]	Perm: Processing provenance and data on the same data model through query rewriting	This paper presents a provenance model similar to that of semi-rings focusing on supporting other operators such as semi-joins.
[74]	Google fusion tables: web-centered data management and collaboration	A presentation of a cloud-based system that facilitates the integration of data on the web. Datasets, e.g. in the form of CSV files or spreadsheets, can be uploaded to the system and made public or shared with collaborators.
[75]	Global detection of complex copying relationships between sources	Methods that are developed to detect copying relationships between sources in order to find the number of independent occurrences of facts are discussed in this paper.
[76]	Crowdsourcing systems on the world-wide web	A survey to get a global picture of crowdsourcing systems on the Web is presented in this paper.
[77]	A Novel Multidimensional Approach to Integrate Big Data in Business Intelligence	In this paper, an approach for integrating different formats into the recent RDF Data Cube format is presented. The approach is based on a MapReduce paradigm.

TABLE 5: The table summarizes and refers to a representative set of papers focusing on Machine Learning and Data Mining in the context of Big Data.

Reference	Title of paper	Description/Objective
[4]	Big data: A survey	This paper discusses technical challenges and reviews advances of the value chain (data generation, data acquisition, data storage, and data analysis) of Big Data
[80]	Data mining with big data	A HACE theorem that characterizes features of the Big Data revolution is presented, and a Big Data processing model, from a Data Mining perspective, is proposed in this paper.
[81]	Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners	A look into how to leverage Big Data analytics efficiently in order to foster positive change.
[82]	Mining Big Data: current status, and forecast to the future	This paper forecasts the future and presents the current status and controversies when it comes to some of the most interesting state-of-the-art topics in Big Data Mining.
[83]	Complex Power System Status Monitoring and Evaluation Using Big Data Platform and Machine Learning Algorithms: A Review and a Case Study	This paper reviews intelligent machine learning methods for complex power systems and key technologies in Big Data management.
[84]	Mining Outlier Data in Mobile Internet-Based Large Real-Time Databases	In this paper a novel mining outlier data method for analysing real-time data features is presented.

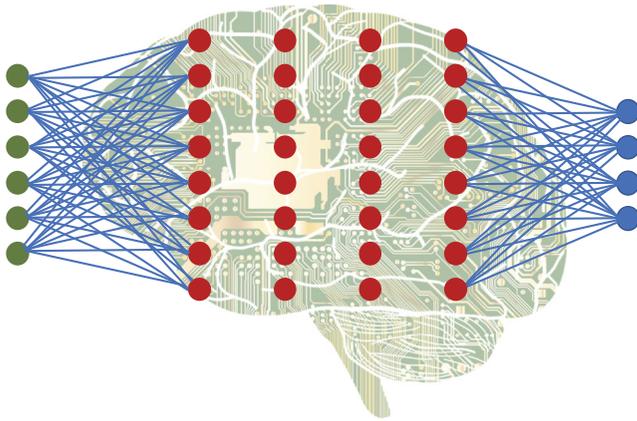


FIGURE 3: As can be seen in the figure, a Deep Learning Neural Network contains several hidden layers. Often the heavy computations are run on GPUs or clusters of GPU servers.

They offer different execution models as standalones or utilize high-performance computing based on, e.g., Hadoop, or Spark Cluster that allows a reduced time of computations. The frameworks have been widely compared and the reviews can be found online (<https://dzone.com/articles/8-best-deep-learning-frameworks>) (<https://www.exastax.com/deep-learning/a-comparison-of-deep-learning-frameworks/>). It should be noticed that these frameworks implement a processing model where the data are transferred to a server performing the analysis and in a final stage the response is returned. This model is subject to latency that could not be acceptable in some applications where there are requirements for high reliability, like, for example, when it comes to autonomous cars [89]. Thus, if efficiency constraints require real-time data processing, then a particular implementation of the algorithm is made on a local node. In its basic setting, this solution does not allow the use of information from other sources. An example of on the node-processing has been presented in [90], where on the node spectral domain preprocessing is used before the data is passed onto the deep learning framework for Human Activity Recognition.

For the IoT the deep analytics are made on large data collections and are usually based on creating more descriptive features of processed objects. For example, in temporal data processing for indoor location prediction [91], a Semisupervised Deep Extreme Learning Machine algorithm has been proposed that improves the localisation performance. The wireless positioning method has been improved with the usage of the Stacked Denoising Autoencoder and that also improves the performance by creating reliable features from a large set of noisy samples [92]. The prediction of home electricity power consumption has been analysed with a deep learning system that automatically extracts features from the captured data and optimises the electricity supply of the smart grid [93].

In Edge Computing with the analytics performed by a deep learning cluster [94], the resource consumption has been efficiently reduced [95]. Convolutional neural networks with automatically created features appeared to be a very

good solution for privacy preservation [96]. Also in the security domain, deep learning finds many applications, e.g., it allows the construction of a model-based attack detection architecture for the IoT for cyber-attack detection in fog-things computing [97].

Video analysis integrated in IoT networks is strongly supported by neural networks, e.g., deep learning-based visual food recognition allows for the construction of a system employing an edge computing-based service for accurate dietary assessment [98]. RTFace, a mechanism for denaturing video streams, has been based on a Deep Neural Network for face detection [99]. It selectively blurs faces and enables privacy management for live video analytics.

3.5. Classification, Prediction, and Visualization. This section discusses the final stage in the chain of the “Procedure for Knowledge Discovery,” which is the obtainment of the final knowledge extracted from the raw data.

When employing machine learning methods for classification and prediction, it is important to use methods with good ability to generalize. The reason for this is that when we apply any of the aforementioned techniques, and after they have been trained on the original data, we want them to make good classifications and predictions of novel data rather than on the data used for training.

After machine learning methods have been applied, it is crucial to know how to interpret their outputs and understand what these mean and how they improve the knowledge in each application area. To that end, visualization methods are employed. Such methods are widely used within Big Data scenarios as they are very helpful for all types of graphical interpretations when the Volume, Variety, or Velocity are complex. In Table 6, we present a summary of, and referral to, papers that deal with visualization.

4. Conclusion

As indicated by the journal articles and the conference papers we have reviewed in this article, the complexity of Big Data is an urgent topic and the awareness of this is growing. Consequently, there is a lot of research carried out on this, and we will in all likelihood find more and more progress in this field during the next few years.

Additionally, a key issue that we really want to emphasize in this study is the aspects related to Big Data which transcend the academic area and that, therefore, are reflected in the company. An observation is that more than 50% out of 560 enterprises thinks Big Data will help them increase their operational efficiency as well as other things [60]. This indicates that there are a lot of opportunities for Big Data. However, it is also clear that there are many challenges in every phase of the knowledge discovery procedure that need to be addressed in order to achieve a continued and successful progress within the field of Big Data.

As is shown in Figure 1, there are three general approaches when carrying out intensive data processing in IoT architectures: (a) local processing, (b) edge computing, and (c) cloud

TABLE 6: The table summarizes and refers to a representative set of papers focusing on Visualization and Prediction in the context of Big Data.

Reference	Title of paper	Description/Objective
[100]	Beyond the hype: Big Data concepts, methods, and analytics	The main objective described in this paper is analytic methods and how are they used for Big Data, in particular, the ones related to unstructured data.
[101]	Key Performance Indicators: Developing, Implementing, and Using Winning KPIs	This book represents a guide with tools and procedures to discover the KPIs and how they are developed and used.
[102]	The visual organization: data visualization, Big Data, and the quest for better decisions	The paper describes data visualization myths, such as: that all data must be visualized, when in fact only good data should be visualized; visualization will always manifest the right decision or action; and that visualization will lead to certainty.
[103]	Big Data and Visualization: Methods, Challenges and Technology Progress	The paper presents applications, technological progress of Big Data visualization, and discusses challenges of it.
[104]	Big-Data Visualization	A special issue which focus on the current situation and new trends of Big Data Visualization.

computing. The text explained each of these approaches more in detail.

We also explained the knowledge discovery procedure by dividing it into several stages as shown in Figure 2. These steps are IoT Data Gathering, Data Cleaning, Integration, Machine Learning, Data Mining, Classification, Prediction, and Visualization.

We have also discussed that many research papers are focused on the variety of information because this is in itself, in conjunction with integration, one of the most challenging issues when it comes to the IoT. This is also the reason why it is very often also associated with one of the most difficult Vs of Big Data, which is the variety of data.

The trend for the future seems to be that more investigations will be carried out in such areas as (a) techniques for data integration, again the V of Variety; (b) more efficient machine learning techniques on big data, such as Deep Learning and frameworks such as Apache's Hadoop and Spark, that will probably have a crucial importance; and (c) the visualization of the data, with, e.g., dashboards, and more efficient techniques for the visualization of indicators.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors acknowledge the support from the research center Internet of Things and People (IOTAP) at Malmö University in Sweden. This work was also supported by the Spanish Research Agency (AEI) and the European Regional Development Fund (ERDF) under project Cloud-Driver4Industry TIN2017-89266-R.

References

- [1] H. Sundmaeker, P. Guillemin, P. Friess, and S. Woelfflé, *Vision and Challenges for Realising the Internet of Things The meaning of things lies not in the things themselves, but in our attitude towards them. Antoine de Saint-Exupéry*, 2010.
- [2] A. Zaslavsky, C. Perera, and D. Georgakopoulos, "Sensing as a Service and Big Data," in *Proceedings of the International Conference on Advances in Cloud Computing (ACC)*, pp. 21–29, 2012.
- [3] L. Atzori, A. Iera, and G. Morabito, "The internet of things: a survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.
- [4] M. Chen, S. Mao, and Y. Liu, "Big data: a survey," *Mobile Networks and Applications*, vol. 19, no. 2, pp. 171–209, 2014.
- [5] S. Sagioglu and D. Sinanc, "Big data: a review," in *Proceedings of the International Conference on Collaboration Technologies and Systems (CTS '13)*, pp. 42–47, May 2013.
- [6] M. Armbrust, A. Fox, R. Griffith et al., "A view of cloud computing," *Communications of the ACM*, vol. 53, no. 4, pp. 50–58, 2010.
- [7] C. Hayashi, *What is Data Science? Fundamental Concepts and a Heuristic Example*, 1998.
- [8] V. Dhar, "Data science and prediction," *Communications of the ACM*, vol. 56, no. 12, pp. 64–73, 2013.
- [9] L. Vangelova, "Data scientist," *Scientific Teaching*, vol. 79, no. 6, pp. 66–67, 2012.
- [10] J. Hardin, R. Hoerl, N. J. Horton et al., "Data science in statistics curricula: preparing students to "think with data";" *The American Statistician*, vol. 69, no. 4, pp. 343–353, 2015.
- [11] T. Hey, S. Tansley, and K. Tolle, *The Fourth Paradigm: Data-Intensive Scientific Discovery*, 2009.
- [12] P. Bevington and D. Robinson, *Data Reduction and Error Analysis for the Physical Sciences*, 1993.
- [13] I. C. L. Ng and S. Y. L. Wakenshaw, "The Internet-of-Things: Review and research directions," *International Journal of Research in Marketing*, vol. 34, no. 1, pp. 3–21, 2017.
- [14] T. Saarikko, U. H. Westergren, and T. Blomquist, "The Internet of Things: Are you ready for what's coming?" *Business Horizons*, vol. 60, no. 5, pp. 667–676, 2017.
- [15] D. Gil, A. Ferrández, H. Mora-Mora, and J. Peral, "Internet of things: a review of surveys based on context aware intelligent services," *Sensors*, vol. 16, no. 7, article 1069, 2016.
- [16] R. Pérez-delHoyo, C. García-Mayor, H. Mora, V. Gilart-Iglesias, and M. D. Andújar-Montoya, "Improving urban accessibility: A methodology for urban dynamics analysis in smart, sustainable and inclusive cities," *International Journal of Sustainable Development and Planning*, vol. 12, no. 3, pp. 357–367, 2017.

- [17] Z. Lv, X. Li, W. Wang, B. Zhang, J. Hu, and S. Feng, "Government affairs service platform for smart city," *Future Generation Computer Systems*, vol. 81, pp. 443–451, 2018.
- [18] J. Macke, R. M. Casagrande, J. A. R. Sarate, and K. A. Silva, "Smart city and quality of life: Citizens' perception in a Brazilian case study," *Journal of Cleaner Production*, vol. 182, pp. 717–726, 2018.
- [19] H. March, "The Smart City and other ICT-led techno-imaginaries: Any room for dialogue with Degrowth?" *Journal of Cleaner Production*, vol. 197, pp. 1694–1703, 2018.
- [20] H. Mora, M. Signes-Pont, D. Gil, and M. Johnsson, "Collaborative Working Architecture for IoT-Based Applications," *Sensors*, vol. 18, no. 6, p. 1676, 2018.
- [21] W. Lee and A. Sharma, "Smart sensing for IoT applications," in *Proceedings of the 13th IEEE International Conference on Solid-State and Integrated Circuit Technology, ICSICT 2016*, pp. 362–364, October 2016.
- [22] Institute of Electrical and Electronics Engineers, *IEEE Std 1451.0™ 2007, IEEE Standard for a Smart Transducer Interface for Sensors and Actuators Common Functions, Communication Protocols, and Transducer Electronic Data Sheet (TEDS) Formats*, 2007.
- [23] T. Islam, S. C. Mukhopadhyay, and N. K. Suryadevara, "Smart Sensors and Internet of Things: A Postgraduate Paper," *IEEE Sensors Journal*, vol. 17, no. 3, pp. 577–584, 2017.
- [24] J. Mendes Jr., M. Vieira, M. Pires, and S. Stevan Jr., "Sensor Fusion and Smart Sensor in Sports and Biomedical Applications," *Sensors*, vol. 16, no. 10, p. 1569, 2016.
- [25] H. Mora, D. Gil, R. M. Terol, J. Azorín, and J. Szymanski, "An IoT-Based Computational Framework for Healthcare Monitoring in Mobile Environments," *Sensors*, vol. 17, no. 10, p. 2302, 2017.
- [26] M. Masoudinejad, A. K. R. Venkatapathy, J. Emmerich, and A. Riesner, *Smart Sensing Devices for Logistics Application*, Springer, Cham, Switzerland, 2017.
- [27] C. Chen, M. Lin, and X. Guo, "High-level modeling and synthesis of smart sensor networks for Industrial Internet of Things," *Computers & Electrical Engineering*, vol. 61, pp. 48–66, 2017.
- [28] R. Y. Zhong, X. Xu, E. Klotz, and S. T. Newman, "Intelligent Manufacturing in the context of industry 4.0: a review," *Engineering Journal*, vol. 3, no. 5, pp. 616–630, 2017.
- [29] Y. Su, X. Meng, Q. Kang, and X. Han, "Dynamic Virtual Network Reconfiguration Method for Hybrid Multiple Failures Based on Weighted Relative Entropy," *Entropy*, vol. 20, no. 9, p. 711, 2018.
- [30] C.-W. Tseng, F.-H. Tseng, Y.-T. Yang, C.-C. Liu, and L.-D. Chou, "Task Scheduling for Edge Computing with Agile VNFs On-Demand Service Model toward 5G and Beyond," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 7802797, 13 pages, 2018.
- [31] A. Yassine, H. Rahimi, and S. Shirmohammadi, "Software defined network traffic measurement: Current trends and challenges," *IEEE Instrumentation & Measurement Magazine*, vol. 18, no. 2, pp. 42–50, 2015.
- [32] H. Tahaei, R. Salleh, S. Khan, R. Izard, K.-K. R. Choo, and N. B. Anuar, "A multi-objective software defined network traffic measurement," *Measurement*, vol. 95, pp. 317–327, 2017.
- [33] X. Wang, C. Xu, G. Zhao, K. Xie, and S. Yu, "Efficient Performance Monitoring for Ubiquitous Virtual Networks Based on Matrix Completion," *IEEE Access*, vol. 6, pp. 14524–14536, 2018.
- [34] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: vision and challenges," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637–646, 2016.
- [35] M. Satyanarayanan, "The emergence of edge computing," *The Computer Journal*, vol. 50, no. 1, pp. 30–39, 2017.
- [36] *Fog Computing and the Internet of Things: Extend the Cloud to Where the Things Are What You Will Learn*, 2015.
- [37] P. Hu, S. Dhelim, H. Ning, and T. Qiu, "Survey on fog computing: architecture, key technologies, applications and open issues," *Journal of Network and Computer Applications*, vol. 98, pp. 27–42, 2017.
- [38] E. Ahmed and M. H. Rehmani, "Mobile Edge Computing: Opportunities, solutions, and challenges," *Future Generation Computer Systems*, vol. 70, pp. 59–63, 2017.
- [39] M. Satyanarayanan, V. Bahl, R. Caceres, and N. Davies, "The Case for VM-based Cloudlets in Mobile Computing," *IEEE Pervasive Computing*, vol. 8, no. 4, pp. 14–23, 2009.
- [40] J. Pan and J. McElhannon, "Future edge cloud and edge computing for internet of things applications," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 439–449, 2018.
- [41] L. J. M. Nieuwenhuis, M. L. Ehrenhard, and L. Prause, "The shift to Cloud Computing: The impact of disruptive technology on the enterprise software business ecosystem," *Technological Forecasting & Social Change*, vol. 129, pp. 308–313, 2018.
- [42] A. Celesti, D. Mulhari, M. Fazio, M. Villari, and A. Puliafito, "Exploring Container Virtualization in IoT Clouds," in *Proceedings of the 2nd IEEE International Conference on Smart Computing, SMARTCOMP 2016*, pp. 1–6, May 2016.
- [43] M. Giacobbe, R. Di Pietro, A. Longo Minnolo, and A. Puliafito, "Evaluating Information Quality in Delivering IoT-as-a-Service," in *Proceedings of the 2018 IEEE International Conference on Smart Computing (SMARTCOMP)*, pp. 405–410, June 2018.
- [44] S. Satpathy, B. Sahoo, and A. K. Turuk, "Sensing and Actuation as a Service Delivery Model in Cloud Edge centric Internet of Things," *Future Generation Computer Systems*, vol. 86, pp. 281–296, 2018.
- [45] R. Lovas, A. Farkas, A. C. Marosi et al., "Orchestrated Platform for Cyber-Physical Systems," *Complexity*, vol. 2018, Article ID 8281079, 16 pages, 2018.
- [46] H. Mora, V. Gilart-Iglesias, R. Pérez-del Hoyo, and M. Andújar-Montoya, "A Comprehensive System for Monitoring Urban Accessibility in Smart Cities," *Sensors*, vol. 17, no. 8, p. 1834, 2017.
- [47] A. M. Osman, "A novel big data analytics framework for smart cities," *Future Generation Computer Systems*, vol. 91, pp. 620–633, 2019.
- [48] J. Santos, B. Volckaert, T. Wauters, and F. de Turck, "Fog Computing: Enabling the Management and Orchestration of Smart City Applications in 5G Networks," *Entropy*, vol. 20, no. 1, p. 4, 2017.
- [49] F. Mora-Gimeno, H. Mora-Mora, D. Marcos-Jorquera, and B. Volckaert, "A Secure Multi-Tier Mobile Edge Computing Model for Data Processing Offloading Based on Degree of Trust," *Sensors*, vol. 18, no. 10, p. 3211, 2018.
- [50] A. Monteriù, M. Prist, E. Frontoni et al., "A Smart Sensing Architecture for Domestic Monitoring: Methodological Approach and Experimental Validation," *Sensors*, vol. 18, no. 7, p. 2310, 2018.
- [51] M. Afrin, M. Razzaque, I. Anjum, M. Hassan, and A. Alamri, "Tradeoff between User Quality-Of-Experience and Service Provider Profit in 5G Cloud Radio Access Network," *Sustainability*, vol. 9, no. 11, p. 2127, 2017.

- [52] M. Yuriyama and T. Kushida, "Sensor-cloud infrastructure—physical sensor management with virtualized sensors on cloud computing," in *Proceedings of the 13th International Conference on Network-Based Information Systems (NBIS '10)*, pp. 1–8, September 2010.
- [53] J. J. Calbimonte, H. Jeung, O. Corcho, and K. Aberer, "Semantic Sensor Data Search in a Large-Scale Federated Sensor Network," *Semantic Sensor Networks*, pp. 14–29, 2011.
- [54] J. Liu, J. Chen, L. Peng, X. Cao, R. Lian, and P. Wang, "An open, flexible and multilevel data storing and processing platform for very large scale sensor network," in *Proceedings of the 2012 14th International Conference on Advanced Communication Technology (ICACT)*, 2012.
- [55] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): a vision, architectural elements, and future directions," *Future Generation Computer Systems*, vol. 29, no. 7, pp. 1645–1660, 2013.
- [56] M. Hassanalieragh, A. Page, T. Soyata et al., "Health monitoring and management using internet-of-things (IoT) sensing with cloud-based processing: opportunities and challenges," in *Proceedings of the IEEE International Conference on Services Computing, SCC 2015*, pp. 285–292, IEEE, July 2015.
- [57] S. Li, L. D. Xu, and X. Wang, "Compressed sensing signal and data acquisition in wireless sensor networks and internet of things," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 4, pp. 2177–2186, 2013.
- [58] H. Mora-Mora, V. Gilart-Iglesias, D. Gil, and A. Sirvent-Llamas, "A computational architecture based on RFID sensors for traceability in smart cities," *Sensors*, vol. 15, no. 6, pp. 13591–13626, 2015.
- [59] F. Liu, Y. Liu, D. Jin, X. Jia, and T. Wang, "Research on Workshop-Based Positioning Technology Based on Internet of Things in Big Data Background," *Complexity*, vol. 2018, Article ID 875460, 11 pages, 2018.
- [60] C. L. P. Chen and C. Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data," *Information Sciences*, vol. 275, pp. 314–347, 2014.
- [61] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. Ullah Khan, "The rise of 'big data' on cloud computing: review and open research issues," *Information Systems*, vol. 47, pp. 98–115, 2015.
- [62] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," *Journal of Big Data*, vol. 2, no. 1, pp. 1–21, 2015.
- [63] S. del Río, V. López, J. M. Benítez, and F. Herrera, "On the use of MapReduce for imbalanced big data using Random Forest," *Information Sciences*, vol. 285, pp. 112–137, 2014.
- [64] A. Halevy, A. Doan, and Z. Ives, *Principles of Data Integration*, Elsevier, 2012.
- [65] A. Y. Halevy, "Answering queries using views: A survey," *The VLDB Journal*, vol. 10, no. 4, pp. 270–294, 2001.
- [66] R. Pottinger and A. Halevy, "MiniCon: A scalable algorithm for answering queries using views," *The VLDB Journal*, vol. 10, no. 2–3, pp. 182–198, 2001.
- [67] M. Brundage, *Xquery: The XML Query Language*, 2004.
- [68] R. Goldman, J. McHugh, and J. Widom, "From semistructured data to XML: Migrating the Lore data model and query language," *Markup Languages: Theory and Practice*, vol. 2, no. 2, pp. 153–163, 1999.
- [69] V. Josifovski, M. Fontoura, and A. Barta, "Querying XML streams," *The VLDB Journal*, vol. 14, no. 2, pp. 197–210, 2005.
- [70] N. F. Noy, "Semantic integration: a survey of ontology-based approaches," *ACM SIGMOD Record*, vol. 33, no. 4, pp. 65–70, 2004.
- [71] A. Doan, J. Madhavan, P. Domingos, and A. Halevy, "Learning to map between ontologies on the semantic web," in *Proceedings of the 11th International Conference on World Wide Web (WWW '02)*, pp. 662–673, ACM, May 2002.
- [72] T. J. Green, "Containment of Conjunctive Queries on Annotated Relations," *Theory of Computing Systems*, vol. 49, no. 2, pp. 429–459, 2011.
- [73] B. Glavic and G. Alonso, "Perm: Processing provenance and data on the same data model through query rewriting," in *Proceedings of the 25th IEEE International Conference on Data Engineering, ICDE 2009*, pp. 174–185, China, April 2009.
- [74] H. Gonzalez, A. Halevy, C. S. Jensen et al., "Google fusion tables: web-centered data management and collaboration," in *Proceedings of the the 1st ACM symposium*, p. 175, June 2010.
- [75] X. L. Dong, L. Berti-Equille, Y. Hu, and D. Srivastava, "Global detection of complex copying relationships between sources," *Proceedings of the VLDB Endowment*, vol. 3, no. 1–2, pp. 1358–1369, 2010.
- [76] A. Doan, R. Ramakrishnan, and A. Y. Halevy, "Crowdsourcing systems on the world-wide web," *Communications of the ACM*, vol. 54, no. 4, pp. 86–96, 2011.
- [77] A. Maté, H. Llorens, E. De Gregorio et al., "A novel multidimensional approach to integrate big data in business intelligence," *Journal of Database Management*, vol. 26, no. 2, pp. 14–31, 2015.
- [78] A. Mukhopadhyay, U. Maulik, S. Bandyopadhyay, and C. A. C. Coello, "A survey of multiobjective evolutionary algorithms for data mining: part I," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 1, pp. 4–19, 2014.
- [79] F. Chen, P. Deng, J. Wan, D. Zhang, A. V. Vasilakos, and X. Rong, "Data mining for the internet of things: Literature review and challenges," *International Journal of Distributed Sensor Networks*, vol. 2015, 2015.
- [80] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, "Data mining with big data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 1, pp. 97–107, 2014.
- [81] J. Dean, *Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners*, John Wiley & Sons, 2014.
- [82] W. Fan and A. Bifet, "Mining big data: current status, and forecast to the future," *ACM SIGKDD Explorations Newsletter*, vol. 14, no. 2, pp. 1–5, 2012.
- [83] Y. Guo, Z. Yang, S. Feng, and J. Hu, "Complex Power System Status Monitoring and Evaluation Using Big Data Platform and Machine Learning Algorithms: A Review and a Case Study," *Complexity*, vol. 2018, Article ID 8496187, 21 pages, 2018.
- [84] X. Liu, Y. Zhou, and X. Chen, "Mining Outlier Data in Mobile Internet-Based Large Real-Time Databases," *Complexity*, vol. 2018, Article ID 9702304, 12 pages, 2018.
- [85] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [86] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent Trends in Deep Learning Based Natural Language Processing," *IEEE Computational Intelligence Magazine*, vol. 13, no. 3, pp. 55–75, 2018.
- [87] D. Yu and L. Deng, "Deep learning and its applications to signal and information processing," *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 145–154, 2011.

- [88] Y. Bengio, "Learning deep architectures for AI," *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–27, 2009.
- [89] Y. Tian, K. Pei, S. Jana, and B. Ray, "Deeptest: Automated testing of deep-neural-network-driven autonomous cars," in *Proceedings of the the 40th International Conference on Software Engineering*, pp. 303–314, May 2018.
- [90] D. Ravi, C. Wong, B. Lo, and G.-Z. Yang, "A Deep Learning Approach to on-Node Sensor Data Analytics for Mobile or Wearable Devices," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 56–64, 2017.
- [91] Y. Gu, Y. Chen, J. Liu, and X. Jiang, "Semi-supervised deep extreme learning machine for Wi-Fi based localization," *Neurocomputing*, vol. 166, pp. 282–293, 2015.
- [92] W. Zhang, K. Liu, W. Zhang, Y. Zhang, and J. Gu, "Deep Neural Networks for wireless localization in indoor and outdoor environments," *Neurocomputing*, vol. 194, pp. 279–287, 2016.
- [93] L. Li, K. Ota, and M. Dong, "When Weather Matters: IoT-Based Electrical Load Forecasting for Smart Grid," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 46–51, 2017.
- [94] H. Li, K. Ota, and M. Dong, "Learning IoT in edge: deep learning for the internet of things with edge computing," *IEEE Network*, vol. 32, no. 1, pp. 96–101, 2018.
- [95] Y. Huang, X. Ma, X. Fan, J. Liu, and W. Gong, "When deep learning meets edge computing," in *Proceedings of the 2017 IEEE 25th International Conference on Network Protocols (ICNP)*, pp. 1–2, October 2017.
- [96] S. Sharma, K. Chen, and A. Sheth, "Towards practical privacy-preserving analytics for IoT and cloud based healthcare systems," *IEEE Internet Computing*, vol. 22, pp. 42–51, 2018.
- [97] A. Abeshu and N. Chilamkurti, "Deep Learning: The Frontier for Distributed Attack Detection in Fog-To-Things Computing," *IEEE Communications Magazine*, vol. 56, no. 2, pp. 169–175, 2018.
- [98] C. Liu, Y. Cao, Y. Luo et al., "A New Deep Learning-Based Food Recognition System for Dietary Assessment on An Edge Computing Service Infrastructure," *IEEE Transactions on Services Computing*, vol. 11, no. 2, pp. 249–261, 2018.
- [99] J. Wang, B. Amos, A. Das, P. Pillai, N. Sadeh, and M. Satyanarayanan, "A scalable and privacy-aware IoT service for live video analytics," in *Proceedings of the 8th ACM Multimedia Systems Conference, MMSys 2017*, pp. 38–49, June 2017.
- [100] A. Gandomi and M. Haider, "Beyond the hype: big data concepts, methods, and analytics," *International Journal of Information Management*, vol. 35, no. 2, pp. 137–144, 2015.
- [101] D. Parmenter, *Key Performance Indicators: Developing, Implementing, And Using Winning KPIs*, 2015.
- [102] P. Simon, *The visual organization: data visualization, Big Data, and the quest for better decisions*, 2014.
- [103] L. Wang, G. Wang, and C. A. Alexander, "Big data and visualization: methods, challenges and technology progress," *Digital Technologies*, vol. 1, no. 1, pp. 33–38, 2015.
- [104] D. Keim, H. Qu, and K. Ma, "Big-Data Visualization," *IEEE Computer Graphics and Applications*, vol. 33, no. 4, pp. 20–21, 2013.

