Big Data Research 10 (2017) 33–43

Contents lists available at ScienceDirect

Big Data Research

www.elsevier.com/locate/bdr

Big Data for Context Aware Computing – Perspectives and Challenges

Kalyan P. Subbu *, Athanasios V. Vasilakos *

ARTICLE INFO

Article history:
Received 28 January 2016
Received in revised form 6 July 2017
Accepted 7 October 2017
Available online 18 October 2017

Keywords:
Big data
Context Awareness
Smartphones
Sensors
Internet of Things
Wireless Sensor Networks

ABSTRACT

Big data has arrived. Myriad applications, systems generate data of humongous volumes, variety and velocity which traditional computing systems and databases are unable to manage. The proliferation of sensors in every possible device is also becoming one of the major generators of Big data. Of particular interest in this article is how context aware computing systems which derive context from data and act accordingly, deal with such huge amounts of data. Big industry players namely Google, Yahoo, and Amazon are already developing context aware applications using user data from emails, chat messages, browsing and shopping histories etc. For instance, Gmail reminds us of our flight schedule by understanding flight booking related content in our emails. Similarly, Amazon understands user preference and recommends items of interest to shop and so on. In this paper, we survey context aware computing systems from a Big data perspective. We first propose a taxonomy of existing work on the basis of sensing platforms and then discuss the latest developments in this field of Big data context aware systems focusing on how such systems deal with various Big data challenges. We conclude the paper with an insight on open research issues involving designing and developing context aware Big data generating systems.

1. Introduction

The recent years has witnessed data flow from a multitude of sources. This led to the Big Data revolution. Big data characteristics include a) volume (amount generated), b) velocity (rate of generation), c) variety (originate from various sources) and d) volatality (uncertain and noisy data). A major source of Big data during the last decade originated from Internet related applications such as user emails, multimedia (audio and video), online transactions, shopping and browsing preferences and the Internet of Things (IoT). Other venues identified include scientific data (astronomy), medical (health data records, patient data), biomedical (DNA sequencing), transportation (traffic, routes, taxi demands), entertainment (Netflix), and recommendation systems (travel, music, shopping).

Another domain of interest in this article is Context aware computing (CAC) which enables applications to have an awareness of the context by making inferences from the data collected and providing smart intelligent services to the user. Data sensed is typically used to extract some information about a context, which can refer to 1) Computing/communication context (network connectivity, communication costs, resource accessibility), 2) User context (user profile, location, activity, social situation, preference), 3) Physical context (lighting, temperature, noise, traffic conditions), or 4) Time context (hour of day, day of week, season, year). Domains such as mobile and pervasive computing are two areas which have seen umpteen number of innovative systems with context awareness. Wireless sensor networks (WSNs) a field under mobile computing paved the way for sensing and computation on battery powered devices of small size. However, WSNs lacked the pervasive/ubiquitous property, meaning facilitating sensing and computation anywhere and everywhere. The development of micro eletro-mechanical sensors (MEMS) sensors in devices particularly the mobile phone started an era of pervasive/mobile sensing which was coupled with unobtrusive, continual, and reliable connectivity. Therefore, in contrast to distributed and mobile computing, pervasive computing extends sensing, computing and communication capabilities to the next level by integrating with devices and users, providing complete transparency. Hence, the presence of pervasive sensing environments and ubiquitous availability of devices (smartphones and wearable devices) with embedded sensors and fast processors, has facilitated CAC to a large extent and is therefore no longer confined to a single individual or location or device [1].

Context aware and Big data systems are intertwined. Both systems have sensors generating large amounts of data, require big and efficient storage spaces and implement data analytics to infer knowledge from the data. The context data that originates from large scale scenarios correlates with Big data in terms of volume (several petabytes), variety (1000s of user data such as blood pres-

* Corresponding authors.
E-mail addresses: kalyan_sasidhar@daiict.ac.in (K.P. Subbu), athanasios.vasilakos@ltu.se (A.V. Vasilakos).

https://doi.org/10.1016/j.bdr.2017.10.002
2214-5796/

sure, pulse rate), velocity (streaming at rates greater than storage rates [2]), volatile (uncertain and noisy sensor data). Value from such data has high impact on many diverse applications. The crucial question that needs to be addressed is what value is extracted from data, how is it extracted and where is the value applied? In other words, there must be a meaningful use of Big data [3]. Consider as an example, a smart city where continuous data is sent by millions of sensors installed at traffic intersections, bus and railway stations, weather stations, and sensors embedded in smartphones and wearable devices. All this data define context in terms of ambient environmental conditions(temperature, humidity); user location, activity; time of the day (bus arrival/departure) and so on. Value needs to be extracted from these streams of data, to provide context aware services for instance, accurately predicting and notifying users about weather and suggesting public mode of transport instead of using personal vehicles. Several examples of CAC applications for smart cities have been proposed in the literature [4,5].

There have been variants of survey articles discussing about Big data. For instance, Tsai et al. [6] discuss different data analytic techniques for analyzing large heterogeneous data including various frameworks and platforms available for Big data. Min et al. [7] provide a detailed survey on Big data, challenges with respect to data representation, data redundancy, data life cycle management, data analytics, data security and energy consumption of systems that process large data. Hu et al. [8] characterize the structural differences between traditional and Big data. They survey existing technologies for data generation, acquisition, processing and storage. Similar survey is presented by Chen et al. [9]. Chen and Lin [10] discuss the techniques available for learning from massive amounts of data. They introduce deep learning which is a paradigm where machine learning techniques automatically perform hierarchical learning of Big data for classification. Lee et al. [11] present challenges of handling Big data in a purely medical set up. The IoT domain involves all possible things and objects provided with sensors and IP connectivity resulting in billions of data packets. Papers [12], [13] and [14] discuss data mining for IoT and resource discovery mechanisms. On the other hand, there have also been exploratory and survey articles on Big sensor data. Ang et al. [15] discuss sensors systems such as air pollution monitoring, transportation, disaster management and assistive living that are deployed on large scale and how valuable information is extracted. Cook et al. [16] discuss the challenges and requirements for deploying pervasive computing applications on a large scale. However, they do not involve the concept of Big data and how CAC relates to Big data.

In contrast, this article will discuss systems that comprise of distributed wireless sensor networks, smartphone and wearable sensors that generate large amounts of data and how this deluge of data needs to be handled through the Big data paradigm. For this review, we analyze, compare, classify CAC systems from a Big data systems perspective and the challenges that need to be addressed. We also propose a taxonomy of context aware systems. The taxonomy is categorized by the medium of sensing such as wireless sensors, wearable sensors and smartphone sensors. Using this taxonomy we discuss a) what kind of context aware systems have been designed and developed, b) what kind of data are sensed and what kind of inferences are made on that data by such systems, c) how and what challenges are addressed and d) what challenges do small scale systems need to address in order to transition to Big data systems.

To summarize, this survey plus discussion paper differs from other articles in the following ways:

1. It culminates context aware systems and Big data applications.
2. Presents a novel categorization of context aware systems on the basis of sensing platforms.
3. It discusses some existing large scale context aware systems and how they handle Big data.
4. Discusses the challenges that existing works have failed to address and puts forth opinions.
5. Finally it presents some perspectives on this new research area.

The road map is as follows: We start with a discussion on context aware computing essentials in Section 2 which includes definitions of context and context awareness, evolution of context, and how challenges in Big data are correlated with context life cycle. Section 3 proposes a taxonomy of CAC systems under the Big data domain, critically discusses the representative works under each domain of the taxonomy. Section 4 puts forth the challenges required to be addressed by existing and future CAC systems while solving Big data problems. This is followed by our brief discussion and perspectives on this joint topic of research. We finally conclude the article in Section 5.

2. Context aware computing

In this section, we present an overview of context, context awareness, and context aware computing. Next, Big data challenges are outlined for each stage in the cycle where data is acquired and processed to derive context.

2.1. What is context?

The term Context has been given many definitions over the years by researchers. A commonly used definition is information that can be used to relate or characterize an entity [17]. Using various definitions and interpretations, we summarize what context means in terms of the source of data we discuss in this article, i.e., a sensor. Context is information extracted from raw sensory data originating from many platforms, the data being user/environment/activity centric.

2.2. What is context awareness?

Context awareness was first introduced by Schilit and Theimer [18] in 1994. With the progress of technology and applications, a number of definitions were given of which one of them by Abowd et al. [17] which says important criterion for a system to be context-aware is if it uses the context to provide relevant information and/or services to the user, depending on the task performed by the user.

Context aware systems can be passive wherein the systems constantly monitor the environment and offer appropriate suggestions to users so they can take actions. For instance, when a user walks into a shopping mall, the mobile phone alerts the user with a list of discounted products to be considered. This involves identifying the location of the user and correlating his location with landmarks such as stores, restaurants etc. Alternately, active systems continuously monitor the situation and act autonomously. For example, the system can monitor, identify an environment and modify the sound profile of a user's mobile phone [19].

2.3. Facilitating context aware computing

The meaning of context has evolved over the past two decades and CAC systems have been developed using variants of contexts, ranging from location to environment to device to user context.

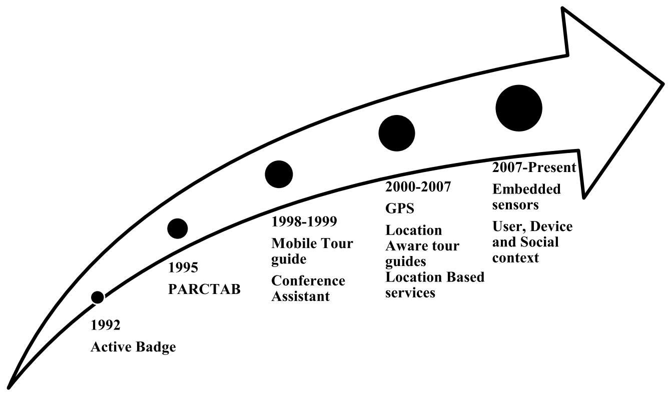

Fig. 1. ActiveBadge, PARCTAB and Mobile tour guides perceived location as context. GPS led to outdoor location based systems and services. Embedded sensors enabled indoor location systems by sensing user activity and ambient environmental conditions. Context further evolved to user's mental, social state and interactions.

- Location of an office, or a user's proximity to a device was considered as context and this context was identified using user worn badges [20], portable devices like palm tablets [21].
- Ambient environmental context includes the profile of an environment based on ambient noise [22], light [23] and other modalities.
- Systems modeling device context include automatically controlling the microphone based on the application, optimizing screen brightness and other applications based on the battery power [24].
- User context includes his/her physical activity, his mode of transport and recently application preferences, usage styles, shopping preferences etc.

Fig. 1 illustrates evolution and growth of context in CAC applications.

2.4. Context life cycle

We briefly describe the stages in the context life cycle which depicts how context moves in context-aware systems starting from raw data, to being extracted as a meaningful context to being used for a meaningful purpose.

- Data acquisition – Acquiring from diverse sensors:
  Sensors are used commonly to measure a physical quantity and convert it into a readable digital signal for processing (and possibly storing). Sensor types include acoustic, sound, vibration, automotive, chemical, electric current, weather, pressure, thermal, proximity, Electrocardiogram (ECG), oxygen saturation (SpO_2), heart rate (HR), blood glucose (BG), respiratory rate (RR), and blood pressure (BP). Fig. 2 depicts the different kinds of sensors that are responsible for generating such data.
- Data awareness – Preprocessing, establishing context:
  Due to the occurrence of noise, motion artifacts, and sensor errors, preprocessing of the raw data is necessary. Usually filtering is performed to remove artifacts or removing high frequency noise. Commonly applied frequency domain techniques are power spectral density (PSD), Fast Fourier transforms (FFT), and low-pass/high-pass filtering. When the data is gathered from numerous wearable sensors, normalization and synchronization of sensor data is also required.
- Data analytics – Making inferences from contextual data:
  Generally, for mining massive and real world data sets, the abstraction of raw data in any data mining approach is a way to design and build a model in order to retrieve valuable information. The aim of feature extraction is to discover the main characteristics of a data set which are identically representatives of the original data [25]. The extracted features should provide meaningful representation of the sensor data which can further help in formulating a relation between raw data and expected knowledge for decision making.
  The features are used by classification algorithms – both supervised and unsupervised to create training data classes and build a classification model, which will then check whether an unknown feature vector (test class) belongs to any class created. Supervised methods include Naive Bayesian and Decision Trees, where labeling of data has to done. In contrast, unsupervised methods like clustering techniques such as k-Nearest Neighbor (k-nn) work on the basis of similarity factors. Unsupervised neural network techniques such as Kohonen Self-Organizing Map (KSOM) are used to classify incoming sensor data in a real-time fashion.

Fig. 3 depicts the context life cycle and the Big data challenges for each stage.

3. Taxonomy of context aware systems

Many taxonomies have been proposed for context aware systems. We mention how the taxonomies are designed and then compare them with our proposed taxonomy.

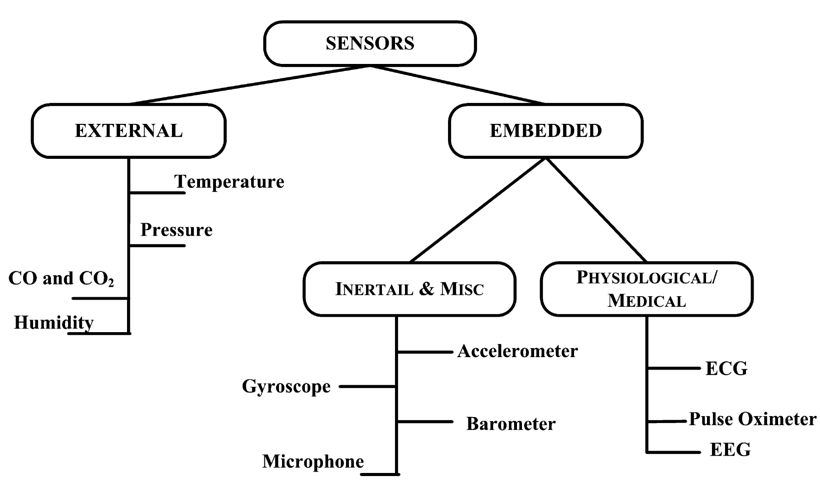

Fig. 2. The figure illustrates various sensors according to their function domain. Embedded sensors in smartphones are inertial type mostly used for sensing motion, angular rotation whereas sensors in wearables can be a mix of inertial as well as sensors that can measure biometric and physiological signals.

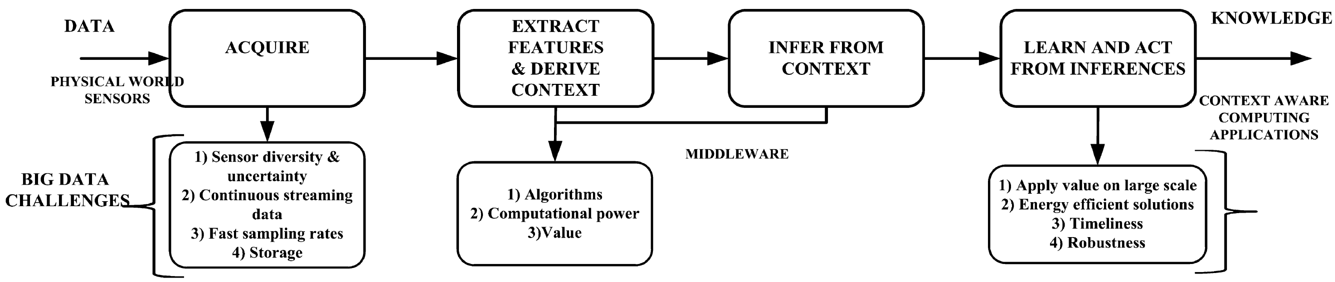

Fig. 3. At each stage of the context life cycle, Big data challenges involve handling/storing heterogeneous data, computational complexity during preprocessing and processing on a large scale, and robust machine learning algorithms for analytics [26].

Baldauf et al. [27] classify context aware systems based on architectures for acquiring context. Rok et al. [28] present a taxonomy of eight categories such as the type of system, application domain, context abstraction, context storage etc. Salkham et al. [29] propose a taxonomy of collaborative context aware systems based on goals, approaches and means. In other words, they differentiate context aware systems based on a) the task taken to achieve or complete, b) the means by which the task is achieved such as via sensing, fusion, actuation, inference etc., and c) the approach whether distributed or collective decision making method. Under these categories fall systems like Smart-Its, Cooperative Artefacts, CoSense, frameworks. In [30], a five layer architecture for context aware systems is proposed as consisting of concept and research layer, network layer, middleware layer, application layer and user infrastructure layer. Works are categorized based on the layered model.

In [31], CAC systems are presented with an Internet of Things (IoT) perspective. They first describe evolution of Internet, basics of IoT and its relation to WSNs. The paper shows how IoT depends and builds upon WSNs by using the hardware (e.g. sensing and communication) infrastructure support to provide access to sensors and actuators. The initial part of the paper concludes with a list of works in the IoT domain which do not have any context aware functionality. The next part of the paper provides detailed description about context awareness and multiple stages of context life cycle based on which their taxonomy is designed. Finally, they discuss and list challenges such as the need for a) automated configuration of sensors i.e., enabling millions of sensors to be automatically connected to applications rather than performing a manual practically cumbersome task; b) proper understanding of data and related context; c) standardization of components and techniques for acquisition, modeling, reasoning, and distribution; d) context sharing between multiple middleware solutions due to the distributed nature of IoT applications and systems; and Security, privacy, and trust. Authors in [32] discuss context awareness in portable sensing systems and review the literature on this topic. The article provides long description very similar to [31] about context awareness fundamentals. The taxonomy is based on application domains like health, transportation, activity recognition, social networking and environmental monitoring. The discussion on each work is brief. No large scale systems and the challenges associated with them are reviewed.

We bring out the main differences between three papers which have a similar theme of context aware computing and Big data. [15] is centered on urban sensing systems that produce large data but no context awareness is mentioned, whereas [31] contains mostly content about context awareness, various context aware frameworks and management systems through an IoT development angle. No working systems or prototypes with end results are discussed. On the other hand [32] talks about existing context aware systems developed on portable platforms but there is no Big data component in the paper. We aim to bridge the gap

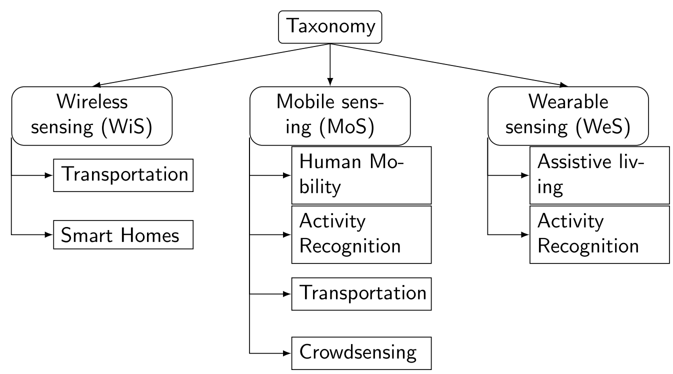

Fig. 4. Each sensing platform has a set of systems developed for certain Big data applications. Activity recognition under MoS refers to transport mode detection i.e., car, bus or bike whereas in WeS it refers to human physical activity like walking, standing, sitting and running. Similarly, using WiS, problem of predicting passenger demands/requests and congestion status are solved under Transportation domain whereas Taxi demand problem is solved using MoS under the same domain.

between these taxonomies by combining Big data, context awareness and different sensing platforms. Our taxonomy is based on the sensing mediums namely Wireless sensing (WiS) through a network of wireless sensing and computing nodes, Mobile sensing (MoS) through sensor rich smartphones and Wearable sensing (WeS) through wearable devices and the different applications that are developed under these categories with a Big data focus.

The reasons for such a taxonomy are multifold:

1. To showcase how CAC systems use one of the sensing mediums to generate Big data and how they handle such data.
2. There is an exponential growth of sensor data from all quarters [33], leading to large scale sensor data analytics [34,35].
3. Smartphones have already seen a tremendous growth in the number of sensors embedded [36].
4. Wearable devices with several sensors including those that can sense sleep patterns, blood pressure, heart rate, sweat levels are commonly seen in fitness and healthcare devices such as FitBit, Apple's and Samsung's smart watches.
5. Application domains of these sensors coincide with those of Big data namely health, environmental, transportation, smart homes and smart buildings.

Fig. 4 illustrates our proposed taxonomy.

We further break down this first level into different domains under which we discuss various systems and applications. We discuss only systems that receive data from sensors and not other means such as Twitter, FB posts, IT services etc. We focus on systems that have a context aware application as an end-to-end solution tied to it rather than a conceptual one. We do not discuss any context aware toolkits, middleware etc.

Table 1
Big data sensing systems: The medium column indicates the sensing medium indicated by WiS, MoS, WeS and H (Hybrid); Context is indicated by HPS for Human physiological signals, SH for Smart Homes; L, D and T for location, day and time of the week; HA for human activity; HM for human mobility; ✓and × indicate whether the systems can or cannot be categorized as Big data and Context Aware; D_s and D_t indicate the data size and data type.

Citation	Medium	Context	BD	CA	D_s	D_t
[37]	WiS	L	✓	×	450 GB	GPS
[38]	WiS	L	✓	×	5 million	GPS
[39]	H	SH	✓	✓	1.5 million	Pressure, Temp, RFID etc.
[40]	WiS	L, D, T	✓	×	100K tuples	Electricity sensor stream
[41]	MoS	HM	✓	✓	130 million location entries	Acc, GPS, CDR
[42]	MoS	HM	✓	✓	35,000 activities	Acc, GPS, CDR
[43]	MoS	HA	✓	✓	38 million	Acc
[44]	MoS	HA	✓	✓	150 hours	Acc
[45]	MoS	HM	✓	×	2.6 million location entries	GPS, CDR
[46]	MoS	L	✓	×	11 million	GPS
[47]	MoS	HM	✓	×	10 users	GPS, CDR
[48]	MoS	User app usage	✓	✓	8 million	App
[49]	MoS	Ambient RF	✓	✓	10 GB	Wi-Fi data
[50]	MoS	HM	✓	✓	1100 users, 17K routes	GPS, CDR
[51]	WeS	HPS	✓	✓	315 TB/year	Acc
[52]	WeS	HPS	✓	✓	35K samples/patient	BP, HR
[53]	WeS	L, HA	✓	×	50K users	Acc, Light, Pressure, Humidity

Table 1 summarizes each of the works we will be discussing. The table shows the functionality of the projects, the size of the data sets used and what kind of context was extracted from the data.

In the next subsection we first provide an overview of each of the sensing mechanisms, then discuss the representative works under each of them. In the last part of each subsection, we list a table that summarizes the Big data related challenges that have been addressed by the proposed systems and the method used to solve.

3.1. Wireless sensing

Wireless sensor networks (WSNs) utilize a wireless network as the means of information transmission. WSNs are preferred when sensing needs to be performed on a phenomenon for which the location is unknown, particularly when the environment to be monitored does not have an infrastructure for either energy or communication. A WSN typically consists of a large number of spatially distributed sensor nodes, which are battery-powered tiny devices. Sensors can be homogeneous or heterogeneous. They are first deployed at random locations to collect sensing data. After sensor deployment is complete, the base station will disseminate the network setup/management and/or collection command messages to all sensor nodes. Based on this indicated information, sensed data are gathered at different sensor nodes and forwarded to the base station for further processing. Mostly WSNs are applied for environmental monitoring and not for context awareness.

Building upon the concept of WSN, we can also consider other external sensors such as GPS chips, sensors that are installed in electrical outlets, water pipes, doors, etc., which have the capability of transmitting sensed data to a central base station server. We will describe wireless sensing systems in Transportation [37,38], and Smart Homes [39,40] where there is an element of context derived and applied unlike traditional WSNs.

3.1.1. Transportation

Smart cities as described through an example in the Introduction section includes the concept of intelligent transportation systems (ITS) which aim at providing smart transport related services based on user location. Examples could be estimating the number of taxis around the user's location and also the time to wait; estimating number of possible passengers requesting rides; prediction of congestion in roads; modeling, predicting and notification of arrivals and departures of public transit systems and so on. In this regards, we discuss two projects that solve the problem of taxi demands and congestion on roads as a Big data problem.

A big data solution for inferring huge number of passenger demands for taxi is proposed in [37]. Taxis enabled with GPS devices are considered as sensors and a network of such wireless sensors is created to collect data and send to a back end office acting as a base station. Traffic records of size close to 450 GB were collected and analyzed for passenger demand. The model developed takes both real-time data uploaded at a particular time and historical data uploaded before that particular time and produces inferred demand in terms of total passenger count for the next slot. Training the complete dataset is performed offline and processing of real-time data is done with a Hadoop cluster with 10 processing nodes.

The authors in [38] propose a Big sensor data system for traffic congestion evolution using deep learning techniques. Deep learning algorithms use multiple-layer architectures to extract inherent features in data at different levels to discover structure in data. 4000 GPS-equipped taxis were used to collect 5 million GPS records containing location, timestamp, and travel speed. These records were used to determine the status of a travel link as congested or not congested. If the computed average speed of a link was lower than a threshold value of (20 km/hour), then the link was marked to be congested. The data is high dimensional time series and therefore they run their deep learning algorithm in a Graphics Processing Unit (GPU) based parallel computing environment to speed up the computation. The proposed model achieved a prediction accuracy of 88% within less than six minutes.

3.1.2. Smart buildings/homes

Smart buildings are buildings retrofitted with wireless sensors and actuators for implementing context awareness. Services of smart buildings include assisted living, monitoring and controlling usage of appliances to save energy, locating people indoors and automatically actuating appliances.

Azzi et al. [39] collect data from infrared sensors, pressure mats, electromagnetic contacts, light and temperature sensors, and radio frequency identification (RFID) tags installed in different equipments in a laboratory smart home set up. Activities such as preparation of coffee, tea, chocolate, cereal, and spaghetti were performed and corresponding data collected. They applied very fast decision tree (VFDT) and obtained classification accuracies of over 95%. The authors addressed big data challenges like handling and processing huge volumes of streaming data by parallel processing using WEKA software a classification tool developed in JAVA. However, no implementation or discussion is presented on handling such heterogeneous, unstructured, noisy and redundant data.

Table 2
Comparison of wireless sensing projects w.r.t. Big data properties and challenges.

Reference	Property satisfied	Challenges addressed	Method
[37]	Vo	Storage & Processing	Hadoop cluster
[38]	Vo, Ve	Storage & Processing	Parallel computing
[39]	Vo, Va	Storage & Processing	Parallel computing
[40]	Vo, Va	Storage & Processing	Hadoop

Hayes et al. [40] propose an anomaly detection system for big sensor data using the context of the data rather than only the content. For instance, detecting abnormally high values of a sensor data when combined with the context (say some activity of interest), then the anomaly detection could result in lower number of false positives. They perform multivariate clustering to group sensors with similar context. More than 100,000 records of sensor data from electrical, water, and gas systems are collected and a MapReduce parallel computing platform is used for offline training of the data. The testing phase is performed real time for detecting anomalies. Results showed that for one subset of data, the number of false positives reduced from 23 to 3 in detecting anomalies, the reason being usage of context along with the content.

3.1.3. Comparisons

We now compare the features of each of the works discussed with respect to Big data challenges. The columns of Table 2 show Reference, which property of Big data did the data collected in the works satisfy (any or a combination of the 4Vs), the challenges addressed and the method used for addressing. We denote Vo for Volume, Ve for Velocity; Va for Variety and Vl for Volatility. We indicate with a — when information is not found. Similar notations are followed in Tables 3 and 4.

3.2. Mobile sensing

The advent of smartphones and wearables with multiple embedded sensors listed in Section 2 are being applied for varied applications. The scalability of smartphone sensing is much easier than its predecessor WSN through a new paradigm Crowdsensing [54], where geographically spanned users actively or passively (opportunistic) involve in data collection with their smartphones and help in building databases of location (GPS, WLAN), motion (accelerometer), proximity (Bluetooth), communication (phone call and SMS logs), multimedia (camera, media player), and application usage (user-downloaded applications in addition to system ones) and audio environment that could be used for research purposes [55].

Crowdsensing has been applied for location analytics and recommendation systems, contextual application usage of large number of users etc. We now discuss implementations of the smartphone platform on large scale in activity recognition [41,56,42], transportation [45,46], environmental [47], indoor location systems [49], recommendation systems [50], mining large scale patterns [57] and two Big data systems [58,59] that could have context awareness integrated in them.

Table 4
Comparison of wearable sensing projects w.r.t. Big data properties and challenges.

Reference	Property satisfied	Challenges addressed	Method
[51]	Vo, Va, Ve	Storage & Processing	Hadoop cluster
[52]	Vo, Va, Ve	–	–
[53]	Vo, Va	–	–

3.2.1. Human mobility prediction

Here, we discuss an interesting article which uses human mobility to understand the most probable activity associated with a specific area in a region.

Santi et al. [41] characterize the mobility not by geographic location but its associated spatial profile. This spatial profile-based mobility pattern, in turn, becomes a human activity pattern. Anonymous mobile phone data, locations are estimated in (latitude, longitude) through cellular network. The location traces were segmented into trajectories to understand daily mobility patterns of each individual. Four different human activities performed on daily basis – eating, shopping, entertainment, and recreational are considered and several points of interests (POIs) for each of these activities are profiled as cells on a map. Then k-means clustering is performed to classify the cells to the most probable activity category. Next, the daily activity pattern for each user is inferred by segmenting an entire day's of activity (24-hours) into eight 3-hour segments, thereby obtaining a sequence of activities performed by the user throughout the day. Using the activity map and user activity patterns, the authors were able to correlate daily activity patterns within a group of people who share a common work area's profile.

3.2.2. Activity recognition

Activity is the most common phenomenon that smartphones have been used for sensing and detection. Activity can be interpreted as physical activity (walking, standing, running, taking stairs, elevators) or the mode of transport (car, bus, train). The accelerometer in smartphones is used to detect such activities. A sample data from an accelerometer collected while a person is walking, standing, moving etc., is shown in Fig. 5.

- Human activity recognition:
  Abdullah et al. [42] propose a framework for handling diverse data collected from a population of 200 users and consists of over 35,000 activities for activity recognition. Scalability is a big challenge in big data systems. This work addresses the challenge by creating separate groups of users based on similarities and then train the data on these clusters.
  Mobile big data challenges and a case study based solution is presented in [43]. The authors implement deep learning based classification on activity recognition data of 560

Table 3
Comparison of mobile sensing projects w.r.t. Big data properties and challenges.

Reference	Property satisfied	Challenges addressed	Method
[41]	W	HPS	Y
[42]	Vo, Ve, Vl	Storage & Processing	Apache Sparke, Deep learning
[43]	Vo, Ve, Vl	Storage & Processing	Apache Sparke, Deep learning
[44]	Vo	Storage	Cloud server
[45]	W	HPS	Y
[46]	Vo	Storage	Cloud server
[47]	Vo	–	–
[49]	Vo	–	–
[50]	Vo	–	–
[48]	Vo	User app usage	Y

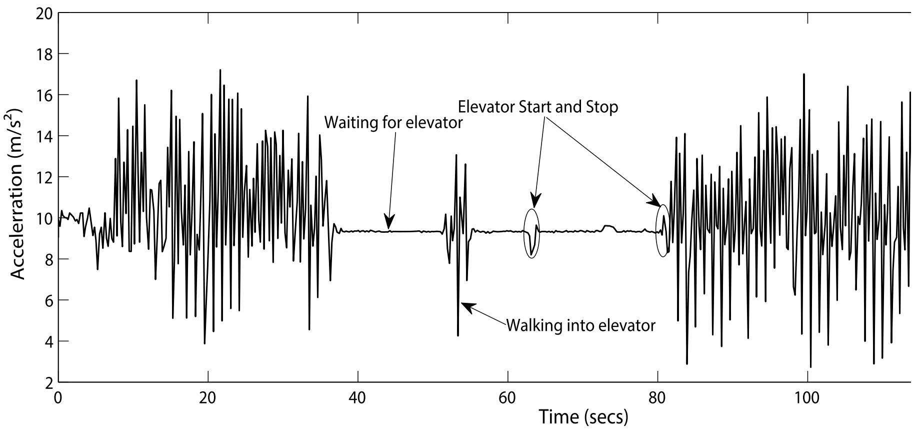

Fig. 5. The accelerometer detects vibrations through the force exerted on its 3-axes (X, Y and Z with Z being the gravity axis). When a person walks, the periodic movement of the leg causes vibrations in the accelerometer, resulting in forces exerted on the axes indicated by the peaks and the small blips indicate jerks caused in the Z axis while traveling in an elevator due the movement against gravity. Features such as mean, median, mode, frequency, power and spectral characteristics from such signals are extracted to derive and define these activities as contexts.

users collected from smartphones. Apache Sparke, an open source framework for scalable MapReduce computing on clusters is used for parallel computing. The framework tackles the volume by parallelizing the learning task into many subtasks each performed on a small partition; velocity by adding streaming extensions for fast and high-throughput processing of streaming data; and volatility by improving the speed up of training deep learning models to update to changes in the data or dynamics of the system.

- Transport mode detection:
  Authors in [44] solving the transport mode detection problem by developing a smartphone application consisting of three levels of classifiers. The first level checks for any user movement. If no movement, the next stage classifies between stationary and moving. Once classifier detects a person to be moving, the last stage classifies the movement as belonging to either a car, bus, train or a tram. The authors propose short data windows to estimate the gravity component by looking for periods where the variation in sensor measurements is small i.e., below a suitable threshold. This indicates the device to be stationary and the force exerted is only due to gravity and not any other component(similar to placing the device on a table with only gravitational forces acting on it). Even on uneven roads, when measurements contain large variation for a long period of time, the gravity component is estimated by dynamically adjusting the variance threshold according to the current movement patterns. Various features mentioned earlier are extracted. The classification is performed on 150 hours of transport data. Results showed 80% accuracy in distinguishing between different transportation modalities. Such inferences of the transport mode and location of an individual can be used to provide a personalized suggestions about the environmental conditions, tracking the hazard exposure and environmental impact of one's activities [56].

3.2.3. Transportation

As mentioned earlier, problems in the ITS domain have been solved using the advantages of crowdsensing through smartphone sensors.

Santi et al. [45] use the probabilistic Naive Bayesian classifier with developed error-based learning approach to predict the number of vacant taxis at a given time and location. The data used is location data of taxis collected through drivers' mobile phones. In order to limit the huge amounts of data from creating hurdles in storage and computation, they follow an information theory centric way of finding the adequate amount of historical data for prediction (choosing a 40 day data set over an entire data set based on the distribution). The work was able to predict taxis with an error or 2–3 taxis per region.

In [46], a platform that can automate the process of collecting and aggregating context information on large scale is proposed. The platform consists of a context manager which controls a monitoring service that is created for collecting and storing data locally on portable devices. Depending on the function, the monitoring service is grouped into push and pull service. Push is activated for collecting context and Pull for sending notifications. The collected data from multiple sensors and sources is aggregated using a semantic model. Finally, actions based on the context information are taken using rule based logic (if-then with OR and AND conditions). To validate their platform, an intelligent transportation application is developed to infer taxi demands and congestion on roads. A traffic collector service is created on top of the platform to collect and send GPS coordinates of taxis. This service is run on smartphones and users are provided with traveling speeds on different roads, average traveling times, and itinerary information. This data was used for predicting availability of free taxis or congestion free routes. The evaluation of the system showed that the system can scale well when more number of users write data (send coordinates). They also propose a context aware framework for efficiently storing large quantities of data. The framework builds upon an existing large-scale distributed storage system BlobSeer [60] to intelligently push the actual data on the cloud but runs the context aware application locally.

3.2.4. Environmental

Ear-phone [47] is a context-aware based noise mapping system developed using smartphones. Here, location of the phone inside a bag, hand or purse is interpreted as context. This context is first detected using an accelerometer and proximity sensor. This form of context awareness helps in measuring data only when the phone's location is acceptable for measurements. For example, when a phone is inside a purse or handbag, the microphone could sound muffled when compared to holding in the hand. Five mobile phones with different hardware are chosen for implementation. Measurement accuracy is one factor that needs to be take care while collecting such data since differences in hardware could result in large variations of data. The authors use one set of data (i.e., from hand) and calibrate the data from other locations like the bag and purse. This is however a heuristic way of handling the variations issue which may or may not suffice for a larger number of phones. The k-nearest neighbor (k-nn) algorithm was used for context classification. Accuracy of around 77% was obtained using data from ten individuals. Scaling the number of users or devices and performing classification could change the result because of factors like sensor diversity, more amount of noisy data, and redundant data.

3.2.5. Location analytics & recommendation systems

Location systems, both outdoors and indoors can be categorized as context aware systems if they have the added functionality of providing smart services based on user location. Outdoor localization requires the GPS in smartphones to be turned ON for activating different services. Faster battery consumption of GPS is limiting factor of outdoor context aware services. On the other hand, there have been many energy efficient solutions for indoor localization using smartphones [61–63]. One large scale implementation of an indoor localization system using different location estimation algorithms on many smartphones was attempted in [49]. The paper proposes a testbed of smartphones for collecting received signal strength index (RSSI) for a large number of locations and implement various estimation algorithms like K-Nearest Neighbor (KNN), Maximum Aposteriori (MAP) and Minimum Mean Square Error (MMSE). Over 1500 locations were covered to fingerprint them with Wi-Fi data. The average estimation error of each algorithm on these location data was calculated and MMSE algorithm resulted in the lowest error.

Continuing on these lines, context aware recommendation systems (CARS) are built upon location analytics. For instance, context aware services like notifying a visitor of a particular store based on his current location and his preferences are being looked at. Intel [50] designed and developed an Apache Hadoop based CARS by building a navigation application, learning the routes a person travels and sends coupons from selected POIs based on contextual factors like location, coupon values, time of day, and weather conditions. To handle the volume of data, the routes were partitioned into regions thereby limiting the number of customers, and also the number of POIs and relevant coupons. The data (which was structured) was preprocessed and analyzed offline on parallel computing infrastructure. The online process consisted of creating a final recommendation list of coupons, storing data on a distributed storage system (not only SQL [NoSQL]). Validation of the system was performed using 1100 customers who performed 17,000 navigations. The metric chosen was conversion rate i.e., how many coupons did customers actually use out of a total number of coupons provided to them. Intel's CARS resulted in a 45% increase in the conversion rate when compared with a system without any contextual intelligence.

Detailed description of other CARS such as travel guides and music recommenders are provided in [64] for further reference.

3.2.6. Contextual application usage

Smartphones are being developed to understand which mobile applications are used by people based on a specific context through users' personal preferences, visiting places of interests, and social interaction levels.

In [57], physical proximity (social context) and semantic places (location context) are considered as two contextual entities. Outdoor locations are sensed through GPS, and indoor locations via the presence of WiFi access points (APs). A list of frequently scanned APs are maintained to denote regularly visited indoor locations. Social context or proximity of user to other users is recorded through the number of visible Bluetooth devices in the vicinity of the phone. A location was considered as stay point if a user stayed for a long time. Multiple stay points were grouped as meaningful locations using a clustering algorithm. Physical locations were manually annotated with a label to create semantic location names like home, office, vacation etc. Regularly used applications such as SMS, Voice, Chat, Calendar, Maps, etc., were chosen to understand which applications were used most based on a particular location. Data analysis was performed on 8.6 million location entries, 6.2 million of Bluetooth readings and 7156 discovered places. The value extracted from data revealed high correlation between number of times an application was used with a place, for instance users used voice call less frequently while at work. Authors suggest that this contextual usage of applications can be used to design interactive user interfaces and also provide related content to the user. Although a valid work in Big data CA systems, no information is provided about the processing and computation platform that was used for such large data.

3.2.7. Non context aware systems

We briefly mention few systems which use crowdsensing to collect and analyze data but do not have a context aware functionality.

Overeem et al. [58] crowdsourced 200 million data sets of different smartphone battery temperature readings and used heat transfer models to estimate the daily air temperatures. Perhaps, context awareness can be incorporated by using the estimated air temperatures in the development of smart cities or alerting the user about the effects of harsh temperature during his/her travel.

Weppner et al. [59], in their work on Bluetooth based crowd counting for effective crowd management, placed Bluetooth sensors near the entrance of a convention center and collected 10,000 Bluetooth devices. Using the data, the authors correlated the crowd movement with a particular event. In other words, the more interesting an event, the more the number of people that gathered. This type of context information could very well be useful in managing large crowds [65].

3.2.8. Comparisons

Table 3 provides a similar comparison to Table 2.

3.3. Wearable sensing

Wearable sensing systems are small form factor devices that can worn a wrist, hip, arms, and neck. The advantage with these systems is the feasibility of designing and interfacing various other sensors such as air quality, pressure, pulse oximetry etc., that are not currently embedded in smartphones. The design and production of wearables is growing exponentially and has taken tremendous strides. Recent proliferation of Google Glass, Smart watches etc., are an addition to the growing list of wearable devices. Wearable sensors are predominantly being applied in health domain [66] particularly assisted living, activity recognition [67], diagnosis and prognosis systems [68,69]. So, we limit our discussion to studies and implementations to Big data wearable health-care systems. Proposals from the literature in health [51,52], HAR [43,53] will be discussed next.

3.3.1. Assisted living

Jiang et al. [51] propose a Big data health care system comprising a wearable sensor, mobile phone and a centralized Big data server. The wearable sensor measures physiological signals, human activity data and transmits them to a mobile phone, which detects the state of the user using Markov models and makes intelligent decisions regarding when to forward the data to the Big data server. The intelligent forwarders provide the remote wearable sensors with necessary context-awareness so that the sensors transmit only important information to the server for analytics when certain behaviors occur and avoid overwhelming communication and data storage. The data generated includes accelerometer, skin temperature, RSSI, heartbeat and SpO_2 and ambient temperature. For a system to support 10,000 users and with data redundancy, the authors estimate a daily raw data consumption of 864 GB and the consumption for one year to reach 315 Terabytes(TB). They also estimate that 222 nodes would be required in the cluster of servers. Although the paper first estimates the size of the data to TB, the implementation is performed on a very small scale with three users.

Forkan et al. [52] develop a Big data framework that analyses medical data of patients and create large scale context awareness for elderly in an assisted living environment. The framework consists of different modules with operations such as data collection and context awareness running in a personal cloud whereas context processing, knowledge acquisition data mining etc., are run in distributed cloud environments. The context states are defined by the activity a person is performing and his vital signs parameters comprising of blood pressure and heart rate. Context awareness is facilitated by learning the parameters of a user and detecting sudden abnormalities in the parameters and sending warning signals. Different classifiers run on the cloud resulted in accuracies of above 85%. Although this work works on data volume wise, it is not clear how challenges such as noisy or redundant data are handled.

3.3.2. Activity recognition

Authors in [53] design and develop a wearable device with an accelerometer, light, pressure and humidity sensor. The authors addressed challenges such as sensor cost, battery and sensor lifetime in the field, adaptiveness and robustness to adverse environmental conditions, by designing a small form factor device with a battery life of seven days and easy to use interface. A cloud component was incorporated with a low delay upload (close to 12 hours delay in typical settings) and scalability feature to serve between 1 and 50,000 users while being cost-efficient. Experiments were conducted to a) measure and learn about daily activities (estimate daily steps taken), b) identify the location of the sensors as being indoors or outdoors based on light, humidity, and activity data and c) identify the travel mode to determine users' CO_2 footprint. From the implementation, it was validated that the platform efficiently accepted sensor data from over 40,000 users and inferences such as activity recognition, and context detection (indoors vs outdoors) were classified with high accuracy.

3.3.3. Comparisons

In Table 4, we were unable to find information about which challenges were addressed and how since certain papers only described the data collected and the implementation.

In the next section, list some challenges that need to be addressed while designing and developing CA systems with Big data features.

4. Discussions

The previous section presented a map of existing context aware systems that handled Big data in different ways. Some proposed ways to store and process, some used the cloud for storing and performed local parallel processing, some implemented context awareness in decided when and how much amount of data to send and so on. In this section, we discuss some challenges in the design and implementation of Big Data context aware systems.

A. Energy consumption:
Wearable sensors due to their proximity to the user sense user related data at a micro level. However, to facilitate this, the sensing units should be continuously running. This could drain the battery of the entire unit. Also these devices should be able to efficiently send the data collected to a nearby server or a mobile phone which can further transfer the data to a server. This requires technologies like Bluetooth or Wi-Fi to be always ON in these devices.

B. Embedded sensor diversity:
Smartphone manufacturers use sensors from different vendors. This leads to a diversity in sensor and when utilized for sensing applications, presents itself a huge challenge. When sensing applications are tested with one phone, care has to be taken to make sure that the application works with same sensors but with different characteristics. This diversity when handled on a large scale will result in Big heterogeneous data. For instance, the Big mobile data repository [55] consists of data sets obtained from different users but all of them had the same phone and same data plan. In a real practical scenario, there is bound to be diversity in the brand of phones and most importantly the sensors, leading to thousands of different phones with different sensors and features which could present an overall different picture.

C. Data heterogeneity:
As seen from the discussions in Section 2, various projects handled data of different types which were structured (individual preferences, and demographics)or semi-structured (GPS, sensor data). It was also shown how prototypes addressed this challenge by using distributed storage, and parallel processing tools like Hadoop and Apache. With more and more heterogeneity in data expected, robust evaluations must be done using these tools.

D. Unreliable or Redundant data or loss of data:
Servers accepting sensor data from billions of smartphones, need to have some kind of analytics to detect redundant data and reject it. For instance when hundreds of devices send images of a historical building from the same angle, the server has to be able to identify redundancy in the images, carefully omit them so that it could save huge amounts of storage space. The server should be capable of detecting data incoming from an unreliable source or a hacker and not include it in further processing or analysis. This is because, a large number of sources makes it hard to monitor the quality of a single source. Sensor malfunctioning, malicious participants or change in human behavior when the subjects are aware of the sensors, can lead to erroneous context inference and prediction.
Data loss is also a factor to consider. For example, in the smart cities paradigm under ITS, data loss could mean insufficient or incomplete information being sent to passengers. A recent work has proposed methodologies for handling data loss and increase data validity [70].

E. Local Vs Centralized processing:
Big data context aware applications should be adaptive in their method of processing data. Depending on the application either batch or real time processing should be performed to detect the context and perform large scale inferences. Each of them have their own set of pros and cons. If batch processing is employed, how many batches and how much data needs to be processed without compromising on the quality of the context inference? Too much or too little data can have their own consequences on machine learning algorithms while inferring context. Big data is normally handled through MapReduce [71] and Hadoop and cluster form of processing [72]. A recent survey paper on parallel processing for Big data systems provides assistance in selecting suitable processing systems for specific applications [73].
Despite this growing trend towards cloud computing and recently mobile cloud, certain requirements such as adaptability, resource scarcity, scalability, mobility, frequent network disconnections, and self-awareness have to be met. A new paradigm named Fog computing is being proposed to reduce

the burden of sending all data to a cloud by sending data to the network edge [74].

An interesting idea to explore could be to use multi-core smartphones for parallel computing tasks. The number of cores in these devices range from quad (4) to octa (8) cores. To exploit such computation power, perhaps local parallel processing using a cluster of smartphones can be performed to reduce the occurrence of bottlenecks in the network while uploading large amounts of data.

F. Efficient communication infrastructures:

If huge volumes of data need to be shared across a network for processing, inferences and recommendations, the network infrastructure should be capable of handling such huge data. For instance, videos from millions of users about an event are bound to consume tremendous amounts of bandwidth. Research should be pursued into exploring the methodologies to accept such data. Performance evaluation models should be developed to analyze the behavior of 3G and 4G wireless links when a deluge of data needs to be transmitted. Factors such as packet loss, delay, and congestion need to be critically analyzed.

G. Scalable context inference algorithms:

All context aware systems require some kind of classification technique to work on the data. This is performed either through supervised (classification) or unsupervised (clustering) method. Supervised classification requires labels for data collected. Most existing work has manually labeled data for cross verification and validation of their classification technique. However, as the scale increases it is close to impossible for a manual labeling scheme considering the variety and volume of incoming data [75]. Data from large scale wireless sensor networks for applications such as intruder detection, pest monitoring, disaster management etc., must label the data appropriately for further analysis. Similarly large scale crowdsourced mobile phone data also poses problems such as data labeling, data redundancy etc.

For example, in indoor localization systems to infer the user's location, data from multiple sensors are combined using sensor fusion algorithms such as Kalman and Particle filters [76]. These algorithms basically compute the location of a user at the next time instant based on the user's location in the previous time instant. When context computation scales to an extent where it has to handle millions of user data, the scalability of these algorithms is not evaluated and is a challenge.

H. Collaborative knowledge transfer of Big data:

Mechanisms should be developed or proposed for knowledge transfer in a collaborative environment. For instance, CSIRO Australia [77] has proposed a Big data system in agriculture consisting of data sizes ranging between Terabytes and Petabytes. The data consists of sensor data from vineyards, oyster farms, livestock and soil moisture. Such data are invaluable to not only the locals and country wide but also world wide for researchers and farmers to understand the nuances of the data and the findings or value extracted from the data. For this to happen, there should be a common knowledge sharing mechanism so that similar data from other countries can be stored and used for various purposes with the consent of the organizations or other research groups.

I. Facilitating Big data context awareness for societal benefits:

Research has to be conducted for not only developing large scale systems but the data obtained from such systems should be utilized for solving societal issues. Two important examples include large scale context aware systems in agricultural and health. Specifically, data collected from large scale wireless sensors can be utilized to forecast environmental and crop related problems, learn the crop growth affecting factors from historical data, and develop intelligent algorithms to maximize the yield. The data should also be used to infer and provide farmer or crop centric information. Similarly, context aware health systems should be able to use data from millions of users to a) understand the behavior of individual users, their activities, their health history etc., and b) provide a concise day-to-day health report or statistics that can warn the user of potential health risks. The concept of smart health systems [5] need to be implemented and evaluated for their scalability and expanded accordingly.

5. Conclusion

Big data deluge has started. Numerous articles based their discussion of Big data on applications such as scientific, astronomical, and biomedical. An exhaustive literature survey revealed that only a few of those published articles relate context aware computing to Big data. Hence, motivated by the lack of a detailed article on these two paradigms, we presented a discussion on Big data and context aware computing and pointed out challenges that need to be addressed if existing small scale context aware applications are to designed for large scale. We identified several other interesting applications such as health care systems, indoor localization, agriculture, transportation, environmental etc., that are implemented on a large scale and described their workings. We pointed out challenges such as data diversity, sensor diversity, robust inference algorithms, power consumption, computational delays, etc. We believe this exploratory survey paper will bring out innovative design and development of large scale wireless sensor networks, smartphone systems, wearable devices.

References

[1] P. Mehra, Context-aware computing: beyond search and location-based services, IEEE Internet Comput. 16 (2) (2012) 12–16.
[2] A. Jacobs, The pathologies of big data, Commun. ACM 52 (8) (2009) 36–44.
[3] C. Bizer, P. Boncz, M.L. Brodie, O. Erling, The meaningful use of big data: four perspectives–four challenges, SIGMOD Rec. 40 (4) (2012) 56–60.
[4] R. Kitchin, The real-time city? Big data and smart urbanism, GeoJournal 79 (1) (2014) 1–14.
[5] A. Solanas, C. Patsakis, M. Conti, I. Vlachos, V. Ramos, F. Falcone, O. Postolache, P.A. Pérez-Martínez, R. Di Pietro, D.N. Perrea, et al., Smart health: a context-aware health paradigm within smart cities, IEEE Commun. Mag. 52 (8) (2014) 74–81.
[6] C.-W. Tsai, C.-F. Lai, H.-C. Chao, A. Vasilakos, Big data analytics: A Survey, J. Big Data 2 (1) (2015).
[7] M. Chen, S. Mao, Y. Liu, Big data: A survey, Mob. Netw. Appl. 19 (2) (2014) 171–209.
[8] H. Hu, Y. Wen, T.-S. Chua, X. Li, Toward scalable systems for Big data analytics: A technology tutorial, IEEE Access 2 (2014) 652–687.
[9] C.P. Chen, C.-Y. Zhang, Data-intensive applications, challenges, techniques and technologies: A survey on Big data, Inf. Sci. 275 (2014) 314–347.
[10] X. Chen, X. Lin, Big data deep learning: Challenges and perspectives, IEEE Access 2 (2014) 514–525.
[11] L.C. Ho, Y. Hyung-Jin, Medical Big data: promise and challenges, Kidney Res. Clin. Pract. 36 (2017) 3–11.
[12] Y. Qin, Q.Z. Sheng, N.J. Falkner, S. Dustdar, H. Wang, A.V. Vasilakos, When things matter: A survey on data-centric internet of things, J. Netw. Comput. Appl. 64 (2016) 137–153.
[13] C. Perera, A.V. Vasilakos, A knowledge-based resource discovery for internet of things, Knowl.-Based Syst. 109 (2016) 122–136.
[14] F. Chen, P. Deng, J. Wan, D. Zhang, A. Vasilakos, Data mining for the internet of things: literature review and challenges, Int. J. Distrib. Sens. Netw. 2015 (2015) 12.
[15] L.-M. Ang, K.P. Seng, Big sensor data applications in urban environments, Big Data Res. 4 (2016) 1–12.
[16] D.J. Cook, S.K. Das, Pervasive computing at scale: Transforming the state of the art, Pervasive Mob. Comput. 8 (1) (2012) 22–35.
[17] G.D. Abowd, A.K. Dey, P.J. Brown, N. Davies, M. Smith, P. Steggles, Towards a better understanding of context and context-awareness, in: Handheld and Ubiquitous Computing, 1999, pp. 304–307.
[18] B. Schilit, N. Adams, R. Want, Context-aware computing applications, in: Mobile Computing Systems and Applications, First Workshop on WMCSA 1994, IEEE, 1994, pp. 85–90.

[19] S. Phithakkitnukoon, R. Dantu, Contextalert: context-aware alert mode for a mobile phone, Int. J. Pervasive Comput. Commun. 6 (3) (2010) 1–23.

[20] R. Want, A. Hopper, V. Falcão, J. Gibbons, The active badge location system, ACM Trans. Inf. Syst. 10 (1) (1992) 91–102.

[21] R. Want, B.N. Schilit, N.I. Adams, R. Gold, K. Petersen, D. Goldberg, J.R. Ellis, M. Weiser, An overview of the ParcTab ubiquitous computing experiment, IEEE Pers. Commun. 2 (6) (1995) 28–43.

[22] D. Siewiorek, A. Smailagic, J. Furukawa, A. Krause, N. Moraveji, K. Reiger, J. Shaffer, F.L. Wong, Sensay: A context-aware mobile phone, in: International Symposium on Wearable Computers, 2003.

[23] M. Azizyan, I. Constandache, R. Roy Choudhury, Surroundsense: Mobile phone localization via ambience fingerprinting, in: Proceedings of the 15th Annual International Conference on Mobile Computing and Networking, MobiCom '09, 2009, pp. 261–272.

[24] N. Ravi, J. Scott, L. Han, L. Iftode, Context-aware battery management for mobile phones, in: 2008 Sixth Annual IEEE International Conference on Pervasive Computing and Communications, PerCom, 2008, pp. 224–233.

[25] B.J. Frey, D. Dueck, Clustering by passing messages between data points, Science (2007) 972–976.

[26] P.e. Agrawal, Challenges and opportunities with Big data: a community white paper developed by leading researchers across the united states, http://cra.org/ccc/docs/init/bigdatawhitepaper.pdf, March 2012.

[27] M. Baldauf, S. Dustdar, F. Rosenberg, A survey on context-aware systems, Int. J. Ad Hoc Ubiq. Comput. 2 (4) (2007) 263–277.

[28] R. Zontar, M. Hericko, I. Rozman, Taxonomy of context-aware systems, Elektroteh. Vest./Electrotech. Rev. 79 (1–2) (2012) 41–46.

[29] A. Salkham, R. Cunningham, A. Senart, V. Cahill, A taxonomy of collaborative context-aware systems, in: UMICS'06, 2006, Citeseer.

[30] J.-y. Hong, E.-h. Suh, S.-J. Kim, Context-aware systems: A literature review and classification, Expert Syst. Appl. 36 (4) (2009) 8509–8522.

[31] C. Perera, A. Zaslavsky, P. Christen, D. Georgakopoulos, Context aware computing for the internet of things: A survey, IEEE Commun. Surv. Tutor. 16 (1) (2014) 414–454.

[32] Ã. Yurur, C.H. Liu, Z. Sheng, V.C.M. Leung, W. Moreno, K.K. Leung, Context-awareness for mobile sensing: A survey and future directions, IEEE Commun. Surv. Tutor. 18 (1) (2016) 68–93.

[33] D. Le-Phuoc, J.X. Parreira, M. Hausenblas, Y. Han, M. Hauswirth, Live linked open sensor database, in: Proceedings of the 6th International Conference on Semantic Systems, 2010, pp. 46:1–46:4.

[34] Sensor data analytics, http://www.eetimes.com/author.asp?section_id=36&doc_id=1326715.

[35] How sensors will shape big data, http://dataconomy.com/2015/01/how-sensors-will-shape-big-data-and-the-changing-economy/.

[36] Sensor growth in smartphones, https://www.qualcomm.com/news/snapdragon/2014/04/24/behind-sixth-sense-smartphones-snapdragon-processor-sensor-engine.

[37] D. Zhang, T. He, S. Lin, S. Munir, J.A. Stankovic, Dmodel: online taxicab demand model from big sensor data in a roving sensor network, in: IEEE International Congress on Big Data, IEEE, 2014, pp. 152–159.

[38] X. Ma, H. Yu, Y. Wang, Y. Wang, Large-scale transportation network congestion evolution prediction using deep learning theory, PLoS ONE 10 (3) (2015) e0119044.

[39] S. Azzi, C. Dallaire, A. Bouzouane, B. Bouchard, S. Giroux, Human activity recognition in big data smart home context, in: IEEE International Conference on Big Data, 2014, pp. 1–8.

[40] M.A. Hayes, M.A.M. Capretz, Contextual anomaly detection in Big sensor data, in: IEEE International Congress on Big Data, 2014, pp. 64–71.

[41] S. Phithakkitnukoon, T. Horanont, G. Di Lorenzo, R. Shibasaki, C. Ratti, Activity-aware map: Identifying human daily activity pattern using mobile phone data, in: Human Behavior Understanding, 2010, pp. 14–25.

[42] S. Abdullah, N.D. Lane, T. Choudhury, Towards population scale activity recognition: A framework for handling data diversity., AAAI 2012, 851–857.

[43] M.A. Alsheikh, D. Niyato, S. Lin, H.p. Tan, Z. Han, Mobile big data analytics using deep learning and Apache spark, IEEE Netw. 30 (3) (2016) 22–29.

[44] S. Hemminki, P. Nurmi, S. Tarkoma, Accelerometer-based transportation mode detection on smartphones, in: SenSys, ACM, 2013, p. 13.

[45] S. Phithakkitnukoon, M. Veloso, C. Bento, A. Biderman, C. Ratti, Taxi-aware map: identifying and predicting vacant taxis in the city, in: Ambient Intelligence, 2010, pp. 86–95.

[46] C. Dobre, F. Xhafa, Intelligent services for big data science, Future Gener. Comput. Syst. 37 (2014) 267–281.

[47] R. Rana, C.T. Chou, N. Bulusu, S. Kanhere, W. Hu, Ear-Phone: A context-aware noise mapping using smart phones, Pervasive Mob. Comput. 17 (Part A) (2015) 1–22.

[48] T.-M.-T. Do, D. Gatica-Perez, By their apps you shall understand them: mining large-scale patterns of mobile phone usage, in: Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia, 2010.

[49] G. Larkou, M. Mintzis, P.G. Andreou, A. Konstantinidis, D. Zeinalipour-Yazti, Managing big data experiments on smartphones, Distrib. Parallel Databases (2014) 1–32.

[50] D. Yonatan, O. Razon, Using Apache Hadoop* for Context-Aware Recommender Systems, IT@Intel White Paper, 2014.

[51] P. Jiang, J. Winkley, C. Zhao, R. Munnoch, G. Min, L.T. Yang, An intelligent information forwarder for healthcare big data systems with distributed wearable sensors, IEEE Syst. 10 (3) (2016) 1147–1159.

[52] A. Forkan, I. Khalil, A. Ibaida, Z. Tari, BDCam: Big data for context-aware monitoring – a personalized knowledge discovery framework for assisted healthcare, IEEE Trans. Cloud Comput. 99 (2015), 1–1.

[53] E. Wilhelm, et al., Wearable environmental sensors and infrastructure for mobile large-scale urban deployment, IEEE Sens. J. 16 (22) (2016) 8111–8123.

[54] N.D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, A.T. Campbell, A survey of mobile phone sensing, IEEE Commun. Mag. 48 (9) (2010) 140–150.

[55] J.K. Laurila, D. Gatica-Perez, I. Aad, O. Bornet, T.-M.-T. Do, O. Dousse, J. Eberle, M. Miettinen, et al., The mobile data challenge: big data for mobile computing research, in: IEEE Pervasive Computing, 2012.

[56] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, M. Srivastava, Using mobile phones to determine transportation modes, ACM Trans. Sens. Netw. 6 (2) (2010) 13.

[57] T.M.T. Do, J. Blom, D. Gatica-Perez, Smartphone usage in the wild: a large-scale analysis of applications and context, in: Proceedings of the 13th international conference on multimodal interfaces, 2011, pp. 353–360.

[58] A. Overeem, J.C.R. Robinson, H. Leijnse, G.J. Steeneveld, B.K.P. Horn, R. Uijlenhoet, Crowdsourcing urban air temperatures from smartphone battery temperatures, Geophys. Res. Lett. 40 (15) (2013) 4081–4085.

[59] J. Weppner, P. Lukowicz, Bluetooth based collaborative crowd density estimation with mobile phones, in: IEEE International Conference on Pervasive Computing and Communications, PerCom, 2013, pp. 193–200.

[60] B. Nicolae, G. Antoniu, L. Bougé, Blobseer: how to enable efficient versioning for large object storage under heavy access concurrency, in: Proceedings of the 2009 EDBT/ICDT Workshops, 2009, pp. 18–25.

[61] J. Hightower, G. Borriello, Location systems for ubiquitous computing, Computer 34 (8) (2001) 57–66.

[62] I. Sabek, M. Youssef, A. Vasilakos, Ace: an accurate and efficient multi-entity device-free WLAN localization system, IEEE Trans. Mob. Comput. 14 (2) (2015) 261–273.

[63] K. Subbu, C. Zhang, J. Luo, A. Vasilakos, Analysis and status quo of smartphone-based indoor localization systems, IEEE Wirel. Commun. 21 (4) (2014) 106–112.

[64] G. Adomavicius, B. Mobasher, F. Ricci, A. Tuzhilin, Context-aware recommender systems, AI Mag. 32 (3) (2011) 67–80.

[65] M. Kjaergaard, Studying sensing-based systems: scaling to human crowds in the real world, IEEE Internet Comput. 17 (5) (2013) 80–84.

[66] H. Banaee, M.U. Ahmed, A. Loutfi, Data mining for wearable sensors in health monitoring systems: A review of recent trends and challenges, Sensors 13 (12) (2013) 17472–17500.

[67] O.D. Lara, M.A. Labrador, A survey on human activity recognition using wearable sensors, IEEE Commun. Surv. Tutor. 15 (3) (2013) 1192–1209.

[68] Y.-L. Zheng, X.-R. Ding, C. Poon, B. Lo, H. Zhang, X.-L. Zhou, G.-Z. Yang, N. Zhao, Y.-T. Zhang, Unobtrusive sensing and wearable devices for health informatics, IEEE Trans. Biomed. Eng. 61 (5) (2014) 1538–1554.

[69] J. Chen, K. Kwong, D. Chang, J. Luk, R. Bajcsy, Wearable sensors for reliable fall detection, in: IEEE International Conference of Engineering in Medicine and Biology Society, 2006, pp. 3551–3554.

[70] Y. Liu, X. Weng, J. Wan, X. Yue, H. Song, A.V. Vasilakos, Exploring data validity in transportation systems for smart cities, IEEE Commun. Mag. 55 (5) (2017) 26–33, https://doi.org/10.1109/MCOM.2017.1600240.

[71] J. Dean, S. Ghemawat, Mapreduce: simplified data processing on large clusters, Commun. ACM 51 (1) (2008) 107–113.

[72] S. Berkovich, D. Liao, On clusterization of big data streams, in: Proceedings of the 3rd International Conference on Computing for Geospatial Research and Applications, 2012.

[73] Y. Zhang, T. Cao, S. Li, X. Tian, L. Yuan, H. Jia, A.V. Vasilakos, Parallel processing systems for big data: A survey, Proc. IEEE 104 (11) (2016) 2114–2136.

[74] C. Perera, Y. Qin, J.C. Estrella, S. Reiff-Marganiec, A.V. Vasilakos, Fog computing for sustainable smart cities: A survey, ACM Comput. Surv. 50 (3) (2017) 32:1–32:43.

[75] L. Zhou, S. Pan, J. Wang, A.V. Vasilakos, Machine learning on Big data: opportunities and challenges, Neurocomputing 237 (2017) 350–361.

[76] V. Fox, J. Hightower, L. Liao, D. Schulz, G. Borriello, Bayesian filtering for location estimation, IEEE Pervasive Comput. 2 (3) (2003) 24–33.

[77] Aussie farmers take a byte out of big data, http://csironewsblog.com/2013/06/22/aussie-farmers-take-a-byte-out-of-big-data/.