Research Article

Connecting Big Data and Context Aware Computing for Improving User Activities

Pooja Kuntal Dhule[a], Dr. Pathan Mohd. Shafi [b]

[a]Computer Department, Marathwada Mitra Mandal's College of Engineering, Pune, India
[b] Professor Computer Department, SKNCollege Of Engineering, Pune, India

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 23 May 2021

Abstract: A decade ago, global data began to climb exponentially. It is primarily aggregated through the worldwide web, including social networks, multimedia files, web search requests, text messages, devices and sensors for the Internet of Things. In extracting useful data from such a large volume of data, there are many difficulties. With the emergence of Widespread & Ubiquitous Computing, the term context awareness is becoming popular. Using context data such as physical context, computational context, and user context/tasks, the context-sensitive systems collect data and modify system actions accordingly. The key purpose of this article is to focus on how such large volumes of data are treated by context-sensitive computing systems. In a satellite navigation device, for example, the user's current position is the reference used to change the visualization automatically (for example. map, arrows, direction etc.). From a Big Data perspective, this paper discusses context-aware computing systems and examines numerous big data challenges

Keywords: Context aware computing, big data, machine learning, ubiquitous computing

1. Introduction

Over the last decade, the expansion in the field of information and communication technology (ICT) has given rise to big data applications, primarily demonstrating the enormous amount of data generated or consumed in all areas of technology and industry[1]. It is estimated that the data volume produced by big data applications will increase rapidly over the years.

The big data can be categorized according to three aspects: (a) volume, (b) variety, and (c) velocity [2]. These categories were first introduced by Gartner to describe the elements of big data challenges. At present, immense opportunities are present for mining huge amounts of such data for modern applications in the field of smart cities, smart transportation, ambient assisted living, sensor technology and healthcare monitoring etc. Big data analytics can assist in examining trends, mining patterns, finding hidden information and making well informed business decisions. Another upcoming domain is Context aware computing (CAC) which makes an application to sense context and extract inferences from the data acquired thereby providing the user smart and intelligent insights. Data sensed is typically used to extract some information about a context, which can refer to Computation context, User context, Environmental context and Timing context. Modern applications which are based on ubiquitous and pervasive computing are becoming famous since use of context awareness.

This survey paper highlights on the following topic:

1. Context aware systems, its categorization, methodologies to handle context of data.

2. Big data, its characteristics

3. Context aware and big data systems with respect to varied application areas.

2. Related work

There is a tremendous increase in applications based on big data due to advances in information and communication technology. The data being generated in these applications can originate new concerns like poor data quality [3]. It causes the need for effective quality control techniques. Such type of data analysis may lead to the wrong business decisions. Based on the systematic review [4] it is found that very few researchers suggested quality improvement criteria. One of the solutions for such problems is context awareness. This system uses context information to take an action based on occurrence of a predefined event and at a certain time. The actions are designed based on three forms[6]. These actions can be presentation-based, according to the execution of services per the user context and on tagging sensor data with context information to be processed at a later time. The design of Context aware systems (CAS) is based on the amount of data and the reasoning level for the given application. Some CAS can be based on the system itself, where all CAS tasks are handled locally. Secondly, the CAS design is based on the use of libraries, frameworks, and tool-kits. Finally the CAS can be designed explicitly, which uses context management systems or middleware infrastructure. Personalization is based on user preferences and expectations where CAS can be implemented. The authors in [7] have also provided CAS classification as Self-Managing Context-Aware Systems for optimization, User-Driven Context-Aware Systems for maximization of the external (user) satisfaction and Value-Sensitive Context-Aware Systems focused on public values regarding services.

3. Context aware computing

One of the areas of pervasive (ubiquitous) computing is context-aware computing. Mobility of devices and use of services make context-aware systems a popular research field these days. The Context in the physical world involves a number of important concerns related to the use of data provided by sensors to the context-aware computing platform, which includes: what to sense for the particular type of context, how to acquire the information needed and how to apply reasoning to that information to infer the context of a user [8]. It is highly needed that programs and services should respond according to the user's situation and behave the way she wants these to be, i.e. Services and systems should be more dynamic [9].

Context-Aware Systems (CAS) implement context as a: "person, place or physical or computational object"[5]. This definition clearly identifies an entity as an 'individual', 'computational device', or 'computational object'. According to wiktionary the context is the surroundings, circumstances, environment, background or settings that determine, specify, or clarify the meaning of an event or other occurrence. Context-aware systems are able to adapt their activities based on current context. Context aware systems watch the environment all the time and propose suitable suggestions to users so they can take necessary actions. For example publishing a user's location to appropriate members of a social network, and allowing retailers to publish special offers to potential customers who are near to the retailers [10].

3.1. Characteristics of CAS

The CAS can be classified according to its context, process of data acquisition, communication models, and accessibility of CAS. Table 1 lists down various categories of context.

Table1. Context Categories

Category	**Primary Context**	**Secondary Context**
Active context	The computer itself defines the contextual information. The application changes behavior as per context automatically	The user explicitly defines the contextual information. User information is provided to the application (for example by personalization)
Passive context	The application keeps updating about the change in context to an Passive context. The application keeps updating about the change in context to an interested user and also preserves it for later use.	The applications' context-aware information is notified to the user for explicit permissions.

Different classification schemes are available for CAS, which include the following:

Active context: This type of context influences the dynamic behavior of the application as per the sensed information.

Passive context: This type of context allows the user to decide how to set the dynamic behavior of the application according to the observed context.

CAS can also be based on the following categorization.

Primary context: which are the contexts that current situation of an entity (for example location, time, and activity)

Secondary context: which is the context can be derived by identifying high-level information (for example environment and relationships) with these sub-problems, it was tried to determine how to evaluate the proficiency of the source books used in mathematics lessons and the usefulness level of the LGS mathematics sample questions published by the Ministry of Education.

3.2 Context life cycle

The CAS follows three basic steps for information management as depicted in Fig. 1 context life cycle.

1) Context acquisition- Sensor based. For example sensors for electrocardiogram, heart rate and blood pressure.

2) Context awareness- Preprocessing of data. When the data is gathered from numerous wearable sensors, normalization and synchronization of sensor data is also required.

3) Context analytics- Deriving meaningful information from data.

In the first stage, sensors are used to measure a physical entity in digital form. Due to the occurrence of missing values, noise, and sensor errors, preprocessing of the raw data is essential. As the data is gathered from numerous heterogeneous sensors, normalization and synchronization of sensor data is also required. Then the feature extraction followed by deriving context from the data is desirable. For making inferences from contextual data, a data mining approach is used to retrieve the meaningful information. The extracted features should provide a meaningful representation of the sensor data which can be useful in raw data and expected knowledge for decision making. At each stage of the CAS life cycle, there are big data challenges. Context acquisition will be affected by heterogeneity in sensors, handling continuous streaming of data and storage required. While extracting features and establishing contexts, extraction techniques should be adopted considering the ‘value’ criteria which fine tune the system performance. Once the context is extracted, to infer knowledge from this big data appropriate machine learning, deep learning algorithms need to be selected. This will result in context aware decisions.

In addition to this, CAS design also depends on the special requirements of a system. In CAS design, context data acquisition plays an important role in specifying the architectural model of the system. There are three different approaches on how to acquire context data discussed as follows:

1) Wireless sensing: It uses a wireless network as the means of information transmission (transportation, smart homes etc.) [11]-[12]

2) Mobile Sensing: It is much scalable due to advent in use of smartphones and wearable sensors (human mobility, activity recognition, crowd sourcing etc.) [1][13][14]

3) Wearable sensing has small devices that can be worn on the wrist, hip, arms, and neck (Assisted living, activity recognition etc.) [15][16]

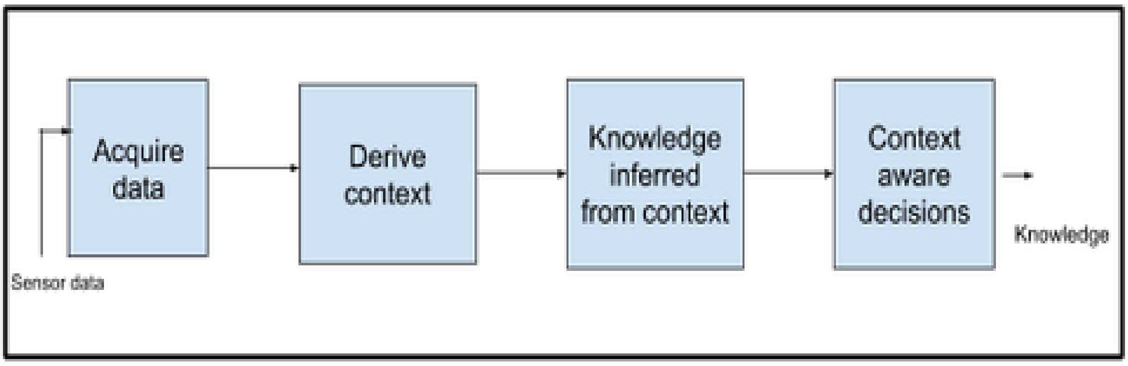

Figure 1 Context life cycle

4. Big data systems

The Big data is characterized in five dimensions as shown in Figure 2 as volume, velocity, variety, veracity and value which corresponds to the amount of data, speed at which data is generated, different forms of data, accuracy or quality of captured data and insights & impact of data. The volume and velocity corresponding to the ‘big word’ which is the base to big data. Variety refers to the many sources and types of data (structured, semi structured and unstructured). Veracity refers to the biases, noise and abnormality in data. both veracity and variety define the ‘data’. The value element depicts the effect of big data on real life applications.

A major source of Big data is World Wide Web related applications such as On-line transactions processing systems, Internet of things, entertainment, recommendation systems-shopping, travel etc., multimedia (audio and video), medical and biomedical. Many researchers studied the ‘big data’ which often represents heterogeneity, availability, complexity which corresponds to the data which cannot be handled by traditional systems for analytical applications [17][18]. Big data demands the use of modern tools, techniques and technologies like classification, clustering, data mining, machine learning, Hadoop, MongoDB etc. This will ensure data analytics

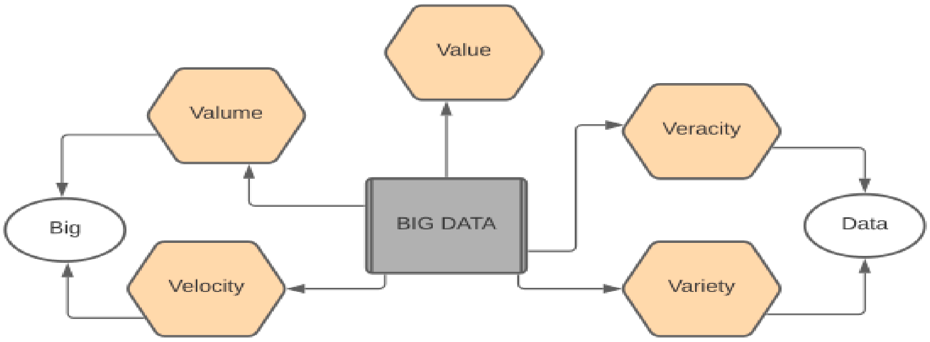

Figure 2 Characterization of big data system

methodologies to extract meaning information from huge amounts of data to ensure enhanced decision making for user activities.

The common risks in such data analytics techniques are information management due to high volume, high variety and high velocity of data. Hence the sophisticated data processing techniques are to be used for discovering new and extracting meaningful knowledge from data. Also to process such a large amount of data the high performance computing is required which will bring out required information from huge amounts of data to facilitate insightful decision making. The common types of big data analytics include predictive, diagnostic, descriptive, and prescriptive analytics. These analytical methods are used to extract diverse knowledge or insights from huge datasets which are to be used at various application domains.

5. Context aware big data systems

Research on big data analytics and context-aware computing has been active for more than three decades, resulting in the development of many concepts, approaches, and systems spanning a large number and variety of application domains. In this section we discuss different representative work based on the mechanism of sensing. Various assisted living applications are based upon use of wearable sensors, which generally measures physiological signals, human activity data and transmits them to a mobile phone and makes intelligent decisions. The contexts are defined by the current activity of a person and his vital signs parameters comprising blood pressure and heart rate. The knowledge of context parameters and sudden change in it could notify abnormalities in values through sending warning signals. Using mobile phone sensors to facilitate transportation, one can detect the mode of transportation whether car, bus or train. One can also predict nearby taxis at a particular time along with GPS information. The user location data is also used to provide personalized recommendations for road traffic congestion, notification of nearby services and map directions etc. Indoor user location aware services are mostly smart home appliances which automate different types of activities with the help of sensors like RFID tags, temperature sensors etc. The human mobility predictions are used to understand the most probable activity associated with a specific area in a region which deals with geographic as well as spatial profile of users. In addition to these application areas stated above the context aware recommendation systems are also based upon location analytics of a user. The user may be notified with the navigation information in a store, a special discount offer etc.

Table 2. Context aware systems applications

Application type	**Authors**	**Year of publication**	**Mode of acquisition**	**Data set and major parameters**	**Context data**	**Methodology**	**Pros**	**Cons**
Assisted Living	P. Jiang et. al.	20 14	Wearable devices, Mobile phone and centralized database server	Live reading from device on wrist. Ambient, skin temperature heart rate	Temperature Acceleration Heart rate	Combined curvilinear distance-based reduction technique and perceptron classifier	Feasibility and interfacing with other devices	Storage and processing of data

	A. Forkan et. al.	20 15	Wearable devices	Physionet MIMIC-II database dataset Systolic BP (SBP) Diastolic BP (DBP) Heart rate	Vital sign: Blood pressure Heart Rate Activity: Current Activity,Last Activity Ambient Conditions	Statistical analysis Correlations Learning and Association Rule Mining-Apriori		
Transportation management	Santi Phithakkitnukoon et. al.	2010	Mobile Sensor data	Taxi-GPS enabled traces for 150 taxis in Lisbon, Portugal	Time, day, week, weather condition.	A na¨ıve Bayesian classifier, probabilistic classifier	Predict vacant taxis at a given time and location	Scalability of data
	C. Dobre a et. al.	2014	Mobile Sensor data	San Francisco taxi dataset	Location user's profile characteristics environment	context ontology, XML rules	traveling speeds on, average traveling times, itinerary information.	
	Samuli Hemminki et. al.	2013	Mobile data	150 hours of transportation data from 16 individuals and 4 different countries	gravity component from accelerometer	kinematic motion,stationary, motorised classifier AdaBoost for prediction	Transportation mode detection	
Recommendations based on user location	D. Yonatan et. al.	2014	Mobile Sensor data, user mobile application usage	1100 customers and 17,000 navigations.	User location and preferences coupon values time of day weather conditions	parallel computing infrastructure Point of interests	Intel's Apache Hadoop	Energy consumption
Smart Building and homes	Azzi, S., Bouzouane et. al.	2014	Wireless sensors and actuators	'SensorIA' of our smart home database	Infrared sensors, pressure mats, light and temperature sensors, RFID tags	Very fast decision tree classifier	Automated activities	Not addressed heterogeneous, unstructured,

	M. A. Hayes et. al.	2014	Wireless sensors and actuators	Electrical, water, and gas systems sensors	electrical, water, and gas systems sensors	Parallel computing platform	Anomaly detection system for big sensor data	noisy and redundant data.
Human mobility prediction & activity recognition	S. Phithakkitnukoon et. al.	2010	Spatial profile based mobility	50,000 mobile phone users	Mobile data location information	Geographic location and associated spatial profile	Activity map and user activity patterns	Scalability of data
	S. Abdullah et. al.	2012	mobile phone data	Public dataset from ALKAN system	Mobile data location information	Similarity network, sensor data, crowd-sourcing	Scalable activity classification	
	Mohammad Abu Alsheikh, et. al.	2016	mobile phone data	Actitracker dataset	Acceleration data	Activities such as jogging, siting, walking, climbing and lying	Spark based deep learning framework	
Environmental	R. Rana et. al.	2015	Accelerometer and proximity sensor.	Live sound recording using n five hardware platforms	Acceleration data, proximity sensor and the light sensor	Mobile phone a MobSLM, sound level meter for measuring environmental sound	k-nearest neighbor (kNN) algorithm	Works when the phone's location is suitable

6. Challenges Need to be Addressed

6.1 Energy consumption

In context awareness, the dynamic context should be evaluated for wireless and wearable sensing related applications such as device monitoring, infrastructure security etc. To sense the required data sensing units should be continuously running. The sensed information should be sent to a nearby server or to a mobile application where the NFC, Bluetooth or wi-fi shall be always ON. Hence designing energy efficient systems is a must.

6.2 Data management strategies to handle erroneous data, data loss and replication

When data is gathered from various sensors, there has to be some mechanism to detect erroneous data. Such erroneous data shall not be considered in analytics further. In object detection when multiple devices capture images of some object, the server must be able to detect erroneous or redundant images. Also there must be some methodology to detect unreliable sources of data. The accurate data analytics will lead to correct inferences and predictions. The insufficient data will also affect context aware applications as incomplete data would be used to model predictions. For example, smart healthcare incomplete, invalid, erroneous data if handled, will result in wrong diagnosis of diseases.

6.3 Efficient use of network infrastructure

The high volumes of data need to be handled for context processing, context inferences and recommendations over the network. The network infrastructure should be sufficient for handling such data. For instance Electronic Healthcare Records (EHR) provides the portability of information between the patient and healthcare practitioners to facilitate the diagnosis of diseases and to understand disease trends. Managing such EHR data and evaluation methodologies are based upon the behavior of network infrastructure that is used to gather such data. Efficient mechanisms are to be developed to handle transmission of data to deal with packet loss, congestion and delay in the network.

6.4 Handling of heterogeneity of network data

The diversity in sensing technology will lead to big heterogeneous data. For instance there are different sensors in a smartphone. The manufacturers use sensors from different vendors. An application when tested on a mobile phone should work with other mobile phones based upon some similar characteristics. This diversity needs to be effectively managed by using various tools like Hadoop for distributed storage and parallel processing.

7. Conclusion

This research has highlighted the concept of context aware and big data systems. This paper presented a discussion on Big data and context aware computing applications such as health care systems, indoor localization, agriculture, transportation, environmental etc. Also the challenges such as data heterogeneity, sensor types, scalable algorithms, efficient use of power etc are discussed. Further, this work can be used for design and development of context aware big data systems.

References

1. Unai Alegre, Juan Carlos Augusto, Tony Clark, Engineering context-aware systems and applications: A survey, Journal of Systems and Software, Volume 117, 2016, Pages 55-83, ISSN 0164-1212, https://doi.org/10.1016/j.jss.2016.02.010.
2. M. Beyer, ‘‘Gartner says solving the ‘Big Data’ challenge involves more than just managing volumes of data,’’ Tech. Rep., AaltoDoc, Aalto Univ., 2011
3. N. Abdullah, S. A. Ismail, S. Sophiayati, and S. M. Sam, “Data quality in big data: A review,” Int. J. Adv. Soft Comput. its Appl., vol. 7, no. Specialissue3, pp. 16–27, 2015
4. M. Mirzaie, B. Behkamal, and S. Paydar, “Big Data Quality: A systematic literature review and future research directions,” Apr. 2019.
5. Abowd, G.D.; Dey, A.K.; Brown, P.J.; Davies, N.; Smith, M.; Steggles, P. Towards a better understanding of context and context-awareness. In Handheld and Ubiquitous Computing, Track 13: e-Health Technologies for Patient Monitoring (ETPM); Springer: New York, NY, USA, 1999; pp. 304–307.
6. Barkhuus L., Dey A. (2003) Is Context-Aware Computing Taking Control away from the User? Three Levels of Interactivity Examined. In: Dey A.K., Schmidt A., McCarthy J.F. (eds) UbiComp 2003: Ubiquitous Computing. UbiComp 2003. Lecture Notes in Computer Science, vol 2864. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-39653-6_12
7. Shishkov B., Larsen J.B., Warnier M., Janssen M. (2018) Three Categories of Context-Aware Systems. In: Shishkov B. (eds) Business Modeling and Software Design. BMSD 2018. Lecture Notes in Business Information Processing, vol 319. Springer, Cham. https://doi.org/10.1007/978-3-319-94214-8_12
8. K.P. Subbu, A.V. Vasilakos, Big Data for Context Aware Computing – Perspectives and Challenges, Big Data Res. (2017), https://doi.org/10.1016/j.bdr.2017.10.002
9. Derrick Ntalasha, Renfa Li and Yongheng Wang, “Internet of Thing Context Awareness Model”, EAI Endorsed Transactions on Context-aware Systems and Applications, Feb. 2016
10. P. S. Gandodhar and S. M. Chaware, "Context Aware Computing Systems: A survey,"2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2018 2nd International Conference on, Palladam, India, 2018, pp. 605-608, doi: 10.1109/I-SMAC.2018.8653786.
11. 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, Towards population scale activity recognition: A framework for handling data diversity., in: AAAI, 2012.
12. 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.
13. S. Abdullah, N. D. Lane, T. Choudhury Omer Berat Sezer , Erdogan Dogdu, and Ahmet Murat Ozbayoglu "'Context-Aware Computing, Learning, and Big Data in Internet of Things: A Survey" IEEE INTERNET OF THINGS JOURNAL, VOL. 5, NO. 1, FEBRUARY 2018

14. M. A. Alsheikh, D. Niyato, S. Lin, H. p. Tan, Z. Han, Mobile big data analytics using deep learning and Apache spark, IEEE Network 30 (3) (2016) 22–29
15. P. Jiang, J. Winkley, C. Zhao, R. Munnoch, G. Min and L. T. Yang (2014) "An Intelligent Information Forwarder for Healthcare Big Data Systems With Distributed Wearable Sensors," in IEEE Systems Journal, vol. 10, no. 3, pp. 1147-1159, Sept. 2016, doi: 10.1109/JSYST.2014.2308324.
16. Forkan, I. Khalil, A. Ibaida, Z. Tari, Bdcam: Big data for context-aware monitoring - a personalized knowledge discovery framework for assisted healthcare, IEEE Transactions on Cloud Computing PP (99) (2015) 1–1.
17. Katal A, Wazid M, Goudar R Big Data: issues, challenges, tools and good practices. In: Proceedings of 6th International Conference on Contemporary Computing (IC3), Noida, August 8–10, IEEE, US, 2013. pp. 404–409.
18. Khan M, Uddin MF, Gupta N. Seven V's of big data understanding: big data to extract value. American society for engineering education (ASEE Zone 1), 2014 Zone 1 conference of the IEEE;