IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

Big Data and Big Data Analytics using Knowledge Discovery Framework in Healthcare System

Sunita Kanadikar [1], Hemanth Kumar N P [2]

M. Tech Student, Dept of Computer Science, Alva's Institute of Engineering & Technology, Shobavana Campus, Mijar, Moodbidri, DK, Karnataka[1]

Assistant Professor, Dept of Computer Science, Alva's Institute of Engineering & Technology, Shobavana Campus, Mijar, Moodbidri, DK, Karnataka[2]

Abstract: Now a days in foreign countries, old age people are adapting the assisted healthcare. An ambient assisted living system (AAL) allows the elderly population to achieve the independency. In this paper we are going to take the use of ambient assisted living system with context aware monitoring to make the context aware decisions for the patient. The Big data collected from the AAL system is analysed in the cloud environment, analysing the trends and pattern with associated probability for individual patient and utilizing this knowledge abnormal condition of the patient can detected and context aware decisions can be made.

Keywords: Cloud, Big data, Assisted HealthCare, Ambient Assisted Living System, Context Aware Decisions, Data Mining.

I. INTRODUCTION

We are staying in the data age and wide variety of data is gathering from the Internet, primary research, location data, social network data, sensor data, device data etc. This kind of data makes the data as big data. Big data is collection of structured, unstructured and semi structured data with different velocity variety volume and veracity. In HealthCare system and mainly in Ambient Assisted Living system the data is coming from the heterogeneous sensors and devices which is generating the large amount of the unstructured raw data for the patient every day. When it comes to Health care the data may be patient profile, medical records, disease history, patient activity, device characteristics, behaviour, routine, temperature, humidity etc. if we store such kind of data for each patient then it leads to necessity of zeta bytes system in few years. So we are suppose go for the cloud based environment for assisted care healthcare infrastructure.

The context aware monitoring requires the extraction of the correct context information, finding the co-relations among the context attributes and predicting the state of the patient by inferred observations and providing the proper services for the patient. The present system for AAL system depends on the Standalone applications with local server or any mobile devices. Such kind of application solve only particular cases and one more problem with the present system is the patient situations are classified by some medical or fuzzy rules, such kind of rules are not applicable for all the kind patient and such kind of system won't sense the future at early stages. In our paper the word context refers to the high level user specific information obtained directly or by sensor devices. These contexts are aggregated and sent to the monitoring canter or doctors. This paper tells one step ahead that is we take all the patient specific data which is already collected and incoming new data are interpreted and decisions can be done. Such kind of Interpretation Helps doctors to make decisions with the greater knowledge. Using this all criteria decisions can be taken about the patient whether he normal, abnormal, alert or emergency such kind of context aware decisions helps to take advance precaution for old age people.

II. ARCHITECTURE

The general architecture of big data analytics has five components like Ambient Assisted living System, Data Collector, Context Aggregate Cloud, Context Provider, and Context Management Cloud. These components are as shown in the Figure 2. The Figure 2 depicts the data is collected from AAL system is sent to Context Aggregator cloud where all context are aggregated to single context. The context management system does the classification applying fuzzy rule generation for the context data.

AMBIENT ASSISTED LIVING SYSTEMS

Ambient assisted living system is mainly designed for elderly population who stays alone and needs the assisted care regularly. The System consists of heterogeneous sensor and different devices which will be surrounded by the patient. Such devices generate huge amount of unstructured data for patient every day. Figure1 shows the ambient assisted living concept which includes the following elements like home security, in community

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

security, in work place. The AAL system consist of low level information consist of raw data collected from the different sources and has the information consist patient health status, location of patient, activities, posture, Motion, emotion, surrounding ambient conditions like (sound, light, humidity, temperature) device status, etc. from these low level context high level context should be obtained for that purpose data should be stored and processed. The Figure 1 shows the AAL system with the big data scenario.

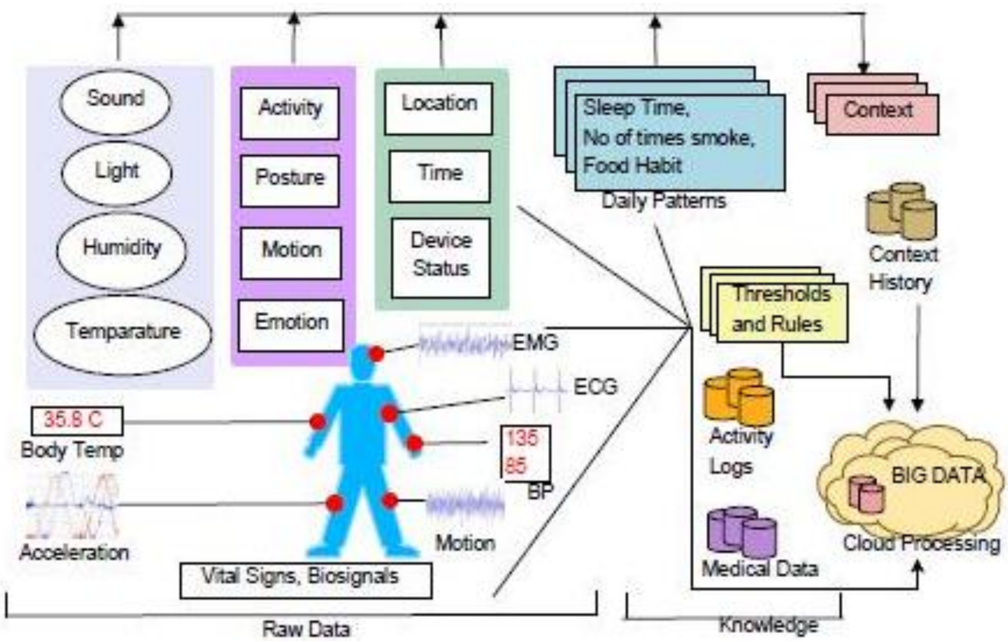

Figure1: Big data Scenario in AAL system

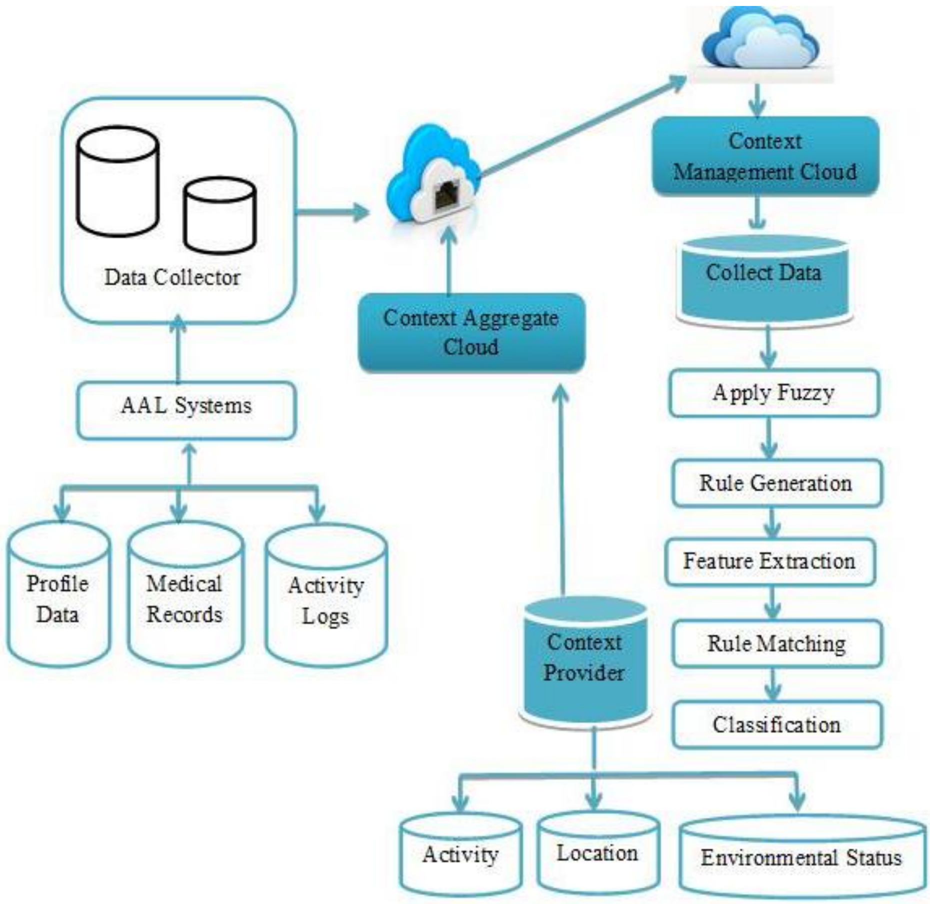

Figure2: Architecture of Big Data Analytics in Health Care System

DATA COLLECTOR

This particular module collect the data from the AAL system and also from the Personal cloud server and gives it to the Context aggregator unit whenever processing is required. The particular unit act like the communication interface, the conversion of high level data to low level data in cloud servers itself

CONTEXT AGGREGATOR CLOUD

Here the Context model will be present what it does is it aggregate all related context to the single context. For example if the patient heart rate is increased its actually the abnormal condition but it checks reasons for increasing the heart rate like if patient is doing exercise then no need to bother

CONTEXT PROVIDER

This particular unit is connected with the context aggregator. The context aggregator connected with multiple number of context provider. Each Context provider applies particular technique to get primitive context from the low level data. For example GPS identifies the location of the user and extract HR value from the ECG data and Accelerometer data helps to identify the current activity of the user. CP converts the possible high level value and gives it to the CA**.**

CONTEXT MANAGEMENT CLOUD

This particular unit is the Central component of the architecture. It connects different clouds so it makes data as big data. If CMS find personalized rules then it will communicate with the PCS and if it finds any new rules then it communicate with the service provider (SP). By communicating with personal cloud server and service provider it keeps the knowledge updated. Then it performs the classification of the situation with the data mining algorithm. When classification is completed it sends the notification to the monitoring centre.

III. SYSTEM FUNCIONALITIES

APPLY FUZZY RULES

Fuzzy rules are like conditional statements which may be in the form of if and then format, for example if the particular condition is satisfies then do the following statement. Consider the example of simple fuzzy rule

If person_height is >180 cm

Then person_weight is > 50kg such kind of rules are used.

RULE GENERATION

Fuzzy rules can be generation depending on two methods first is fuzzy rules based on expert knowledge and second fuzzy rules based on data. In the first rule expert gained knowledge is taken and second method indirectly involves the data gained by the first method.

FEATURE EXTRACTION

The process of feature extraction starts with the raw data and it construct the derived values or features from the collected data depending on the user query.

RULE MATCHING AND CLASSIFICATION

The old rules collected from the system and the new rules collected from the services provider are matched and classification can be done based on the rules.

IV. METHDOLOGY

Two important methodologies are used to perform the system functionalities. To design two algorithms are used first is Map Reduce Apriori it finds the co-relation between the context attribute and map reduce is efficient programming model to process the big data. The second methodology is decision tree algorithms by using this classification decision can be done accurately.

Algorithm1 for context aggregation

1: **Input**: A frequent set of context information C for all AAL systems
2: **Output**: Context state S for each AAL system j
3: **Procedure** Mapper()
4: **begin**
5: **for** each AAL system j **do**
6: **for** domain = 1 to k **do**
7: generate C for time t
8: output(key=(j,t), value=C
9: **end for**
10: **if** c not equal to null **then**
11: output(key=(j,t), value= C)
12: **end if**
13: **end for**
14: **end**
15: **Procedure** Reducer(key=(j,t), value=set of C)
16: **begin**
17: **for** each AAL system j **do**
18: Ci = null
19: **end for**
20: **for** each Cat t in AAL system j **do**
21: Ci=Cj union S
22: **end for**
23: output(key=(j,t), value=Ci
24: **end**

Algorithm1 for classification
Input: Examples, Target attribute, attribute
Output: Decision tree

1: create root node for tree
2: if all the examples are positive, return single node tree root with label = +
3: if all the examples are negative, return single node tree root with label = -
4: if attribute is empty, return single node tree root with label =most common value of the target attribute in example
5: otherwise begin
6: A= the attribute from the Attributes that best classifies the Example
7: The decision attribute for root = A
8: for each possible value vi of A

9: Add new tree branch below the root, A=vi
10: let Examples vi be the subset of Examples that have value vi for A
11: If Example vi is empty
12: then below this new branch add a leaf node with label= most common value of target attribute in Examples
13: else below this new branch add the sub tree and call the function
14: end
15: return root

V. RESULTS

To evaluate the results comparison is done with proposed model and the clinical classification based on some medical rules. The Table 1 shows the comparison it is based on the SBP (systolic Blood pressure) and DBP(diastolic Blood Pressure) and heart rate values by using Generic rules classification can be done only in two ways normal and abnormal for the Big-data model classification can be done on four ways normal, warning, alert, emergency.

Table 1 Comparison of Generic model and Big Data Model

Patient		Generic Rules		Big Data Model			
	Total Data	Normal	Abnormal	Normal	Warning	Alert	Emergency
P1	24789	2	24787	24518	8584	1116	83
P2	23678	8	23670	24911	7980	1360	144

VI. CONCLUSION

The proposed model helps in reducing the false alarms which are happening in the hospitals it helps the care giver to distinguish normal conditions and emergencies. Faster learning with greater knowledge helps doctors to give proper suggestions to patients.

REFERENCES

[1] A. Pantelopoulos and N. Bourbakis, “A survey on wearable sensor-based systems for health monitoring and prognosis,” IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 40, no. 1, pp. 1–12, 2010.
[2] A. K. Dey, “Providing architectural support for buildingcontext-aware applications,” Ph.D. dissertation, Georgia Institute of Technology, 2000.
[3] S. Sridevi, B. Sayantani, K. P. Amutha, C. M. Mohan, and R. Pitchiah, “Context aware health monitoring system,” in Medical Biometrics. Springer, 2010, pp. 249–257.
[4] H. Ding, Y. Moodley, Y. Kanagasingam, and M. Karunanithi, “A mobile-health system to manage chronic obstructive pulmonary disease patients at home,” in Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2012. IEEE, 2012, pp. 2178–2181.
[5] A. Forkan, I. Khalil, and Z. Tari, “Cocamaal: A cloud-oriented context-aware middleware in ambient assisted living,” Future Generation Computer Systems, vol. 35, pp. 114–127, 2014.
[6] R. Buyya, C. Yeo, S. Venugopal, J. Broberg, and I. Brandic, “Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility,” Future Generation computer systems, vol. 25, no. 6, pp. 599–616, 2009.
[7] G. Wu, H. Zhang, M. Qiu, Z. Ming, J. Li, and X. Qin, “A decentralized approach for mining event correlations in distributed system monitoring,” Journal of Parallel and Distributed Computing, 2012.
[8] R. Rastogi and K. Shim, “Mining optimized association rules with categorical and numeric attributes,” IEEE Transactions on Knowledge and Data Engineering, vol. 14, no. 1, pp. 29–50, 2002
[9] Y. Oh, J. Han, and W. Woo, “A context management architecture for large-scale smart environments,” IEEE Communications Magazine, vol. 48, no. 3, pp. 118–126, 2010.
[10] P. Haghighi, A. Zaslavsky, S. Krishnaswamy, and M. Gaber, “Mobile data mining for intelligent healthcare support,” in 42nd Hawaii International Conference on System Sciences, HICSS ’09, 2009.
[11] T. Tamura, I. Mizukura, M. Sekine, and Y. Kimura, “Monitoring and evaluation of blood pressure changes with a home healthcare system,” IEEE Transactions on Information Technology in Biomedicine, vol. 15, no. 4, pp. 602–607, 2011
[12] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, “Data mining with big data,” IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 1, pp. 97–107, 2014.
[13] A. R. M. Forkan, I. Khalil, Z. Tari, S. Foufou, and A. Bouras, “A context-aware approach for long-term behavioural change detection and abnormality prediction in ambient assisted living,”Pattern Recognition, vol. 48, no. 3, pp. 628–641, 2015.
[14] Panorea Gaitanou1,2, Emmanouel Garoufallou3,” The Effectiveness of Big Data in Health Care: A Systematic Review”, Springer International Publishing Switzerland 2014.
[15] J. Cui, C. Li, C. Xing and Y. Zhang, “The framework of a distributed file system for geospatial data management”, Proceedings of IEEE CCIS, **(2011)**, pp. 183-187.
[16] Y. Gang Wang and S. Wang, “Research and Implementation on Spatial Data Storage and Operation Based on Hadoop Platform”, Second IITA International Conference on Geoscience and Remote Sensing, Qingdao China, **(2010)** August 28-31, pp. 275-278.
[17] Y. Zhong, J. Han, T. Zhang and J. Fang, “A distributed geospatial data storage and processing framework for large-scale WebGIS”, 20th International Conference on Geoinformatics (GEOINFORMATICS), Hong Kong China, **(2012)** June 15-17.
[18] Bhandarkar M.: MapReduce programming with Apache Hadoop. In Parallel Distributed Processing (IPDPS), 2010 IEEE International Symposium on, 1–1 (2010).