International Journal on Applications in Information and Communication Engineering
Volume 2: Issue 1: January 2016, pp 30-33. www.aetsjournal.com ISSN (Online) : 2394-6237

A Personalized Knowledge Discovery Framework for Assisted Healthcare using BDCaM

M.Kalpana

Abstract— An ambient assisted living (AAL) system consists of heterogeneous sensors and devices which generate huge amounts of patient-specific unstructured raw data every day. An important feature of remote monitoring applications is to identify the abnormal conditions of a patient accurately and so send appropriate alerts to the care givers. In traditional systems, situations are classified by generalized medical rules or fuzzy rules which are not always applicable for every kind of patient. These systems cannot sense the future at an early stage. In some monitoring systems, when a patient feels unwell he/she needs to press a wearable panic button to notify a response centre about the emergency. In this work, we have presented BDCaM (Big Data – Context Aware Monitoring), a generalized framework for personalized healthcare, which leverages the advantages of context-aware computing, remote-monitoring, cloud computing, machine learning and big data. Our solution provides a systematic approach to support the fast-growing communities of people with chronic illness who live alone and require assisted care. The system can accurately distinguish emergencies from normal conditions. The data used to validate the model are obtained via artificial data generation based on data derived from real patients, preserving the correlation of a patient's vital signs with different activities and symptoms.

Index Terms — Assisted Healthcare. Big Data, Context Aware, Data Mining, Hadoop, Knowledge Discovery, Map Reduce.

I. INTRODUCTION

Context-aware computing is about providing relevent services to users based on situations and "context" is an information that characterizes the situation about the entity of study[9]. Systems take actions independently in Context-aware computing using the model specific to user needs. Context-aware systems find applicability in Healthcare with the objective of improving quality of customer service which is patient care here, by assisting healthcare professionals. Such systems which handle huge volume of data and using technologies such as mobile phone are termed BigData-Context Aware Monitoring systems. These systems are designed to provide real-time healthcare services with the help of state of art technologies [4], [5].

Context-aware monitoring systems adapt themselves based on the data generated in ambient assisted living (AAL) systems [1], [2]. The volume of data generated by AAL systems which constitute a system of sensors and devices may be large [1], even if an AAL system generate few Kilobytes of data every second it will grow to Terabytes in a year [3], [6]. Scaling this to many patients may lead to handling huge volume of data and it is challenging to collect, store, cleanse, analyse and interpret the output. According to Doug Laney, Gartner and IBM data scientists Big Data is characterized by 4Vs – volume, velocity, variety and veracity[10]. BDCaM systems satisfy the above attributes and mining such data may provide useful pattern which are helpful for making decisions by professionals in a cost-effective and timely manner [8]. For example, a selected attribute under monitoring is Blood pressure which has to be in the ideal range of 90 to 120 for adults, if a reading or successive readings are close to the extreme values then professionals can be alerted. Many times this processing may be complex predictive analytics, for example, a combination of multiple attributes may have to be classified using techniques such as decision trees, Support Vector Machine etc. One of the biggest challenge is getting these predictions closer to actuals. For handling such large volume of data one of the most reliable system is Hadoop using Map Reduce programming for processing data.

There have been several studies about the context-aware approach for assisted healthcare. The works are differentiated by: context-aware platforms for supporting continuous care, activity monitoring, cloud-based healthcare, and personalized care[7]. The context-aware systems that are developed using rule-mining and data mining only solve some specific diseases. That is, most proposed systems are restricted to supporting some specific context-aware services and are not capable of detecting a wider range of anomalies. The system that relies on generic rules is not able to predict all the critical situations and suffers from misclassification of normal situations. The studies of big data for health-care mostly focus on the area of mining electronic health records, feature extraction from medical images or pattern recognition based on genome data. A very few works have combined context-awareness with big data to develop a generalized system for assisted care. All these contributions have motivated us to develop this cloud-enabled system with big data. The unique advancement of our model is to learn user-specific anomalies accurately in an assisted living system and take immediate context-aware actions. The robust learning methods reduce

M.Kalpana is currently pursuing M.E Computer Science and Engineering at Pannai College of Engineering affiliated to Anna university, Chennai, Tamil Nadu, India

unnecessary false alerts to the monitoring systems.

HADOOP is a data-intensive cluster computing system, in which incoming jobs are defined based on the Map Reduce programming model. MapReduce is a popular paradigm for performing computations on Big Data in Cloud computing systems A Hadoop system consists of a cluster, which is a group of linked resources. Organizations could use existing resources to build Hadoop clusters - small companies may use their available (he a large company may specify a number of (homogeneous) resources for setting up its Hadoop cluster. There can be a variety of users in a Hadoop system who are differentiated based on features such as priority, usage, guaranteed shares, etc. Similarly, workload in the Hadoop system may have differing numbers of users' jobs and corresponding requirements. Therefore, a Hadoop system can be specified using three main factors: cluster, workload, and user, where each can be either heterogeneous or homogeneous. There is a growing demand to use Hadoop for various applications which leads to sharing a Hadoop cluster between multiple users.

To increase the utilization of a Hadoop cluster, different types of applications may be assigned to one cluster, which leads to increasing the heterogeneity level of workload. However, there are situations where a company assigns a Hadoop cluster to specific jobs as the jobs are critical, confidential, or highly data or computation intensive. Accordingly, the types of applications assigned by different users to a Hadoop cluster define the heterogeneity level of workload and users in the corresponding Hadoop system. Similarly, the types of resources define the heterogeneity of Hadoop clusters.

II. EXISTING AND PROPOSED SYSTEM

A. Hadoop System

Heterogeneity in Hadoop is defined based on the level of heterogeneity in the following Hadoop factors:

Cluster is a group of linked resources, where each resource has a computation unit and a data storage unit. The computation unit consists of a set of slots, where each slot has a given execution rate. In most Hadoop systems, each CPU core is considered as one slot. Similarly, the data storage unit has a given capacity and data retrieval rate. Data in the Hadoop system is organized into files, which are usually large. Each file is split into small pieces, which are called slices. Usually, all slices in a system have the same size.

User submits jobs to the system. Hadoop assigns a priority and a minimum share to each user based on a particular policy (e.g. the pricing policy in the user's minimum share is the minimum number of slots guaranteed for the user at each point in time.

Workload consists of a set of jobs, where each job has a number of map tasks and reduces tasks. A map task performs a process on the slice where the required data for this task is located. A reduce task processes the results of a subset of a job's map tasks. The value defines the mean execution time of job Join resource Investigations on real Hadoop workloads show that it is possible to classify these workloads into classes of "common jobs" We define the class of jobs to be the set of jobs whose mean execution times (on each resource) are in the same range. There are various Hadoop schedulers, where each scheduler may consider different levels of heterogeneity in making scheduling decisions. Moreover, schedulers are differentiated based on different performance metrics (e.g., fairness, minimum share satisfaction, locality, and average completion time) that they address. However, to the best of our knowledge there is no scheduling algorithm which simultaneously considers all of these performance metrics. In some cases, optimizing one metric can result in significant degradation in another metric. For instance, a scheduler which optimizes fairness may need to repeatedly switch the processor between different jobs. This can add significant overhead, which can result in larger average completion times. To analyze the behaviour of schedulers at different levels of heterogeneity, this paper uses three Hadoop scheduling algorithms: FIFO, Fair Sharing, and COSHH. The FIFO and Fair Sharing algorithms are used as the basis of a majority of Hadoop schedulers the COSHH algorithm was first introduced and considers system parameters and state information in making scheduling decisions. These algorithms are selected as representatives of schedulers which consider heterogeneity at different levels. The FIFO scheduler does not consider heterogeneity in its scheduling decisions. However, the Fair Sharing and COSHH algorithms are representatives of schedulers with partial and full consideration of heterogeneity, respectively.

FIFO is the default Hadoop scheduling algorithm. It orders the jobs in a queue based on their arrival times, ignoring any heterogeneity in the system. The experience from deploying Hadoop in large systems shows simple algorithms like FIFO can cause severe performance degradation; particularly in systems that share data among multiple users.

Fair Sharing is a Hadoop scheduler introduced to address the shortcomings of FIFO, when dealing with small jobs and user heterogeneity . This scheduler defines a pool for each user, where each pool has a number of maps and reduces slots on a resource. Each user can use its pool to execute her jobs. If a pool of a user becomes idle, the slots of the pool are divided among other users. This scheduler aims to assign a fair share to users, which means resources are assigned to jobs such that all users receive, on average, an equal share of resources over time. Therefore, the Fair Sharing algorithm only takes user heterogeneity into account.

*COSHH i*s a Hadoop scheduler who considers cluster, workload, and user heterogeneity in making scheduling decisions Using the parameters and state information, COSHH classifies the jobs and finds a matching of the resulting job classes to the resources based on the requirements of the job classes and features of the resources. This algorithm solves a Linear Programming problem (LP) to find an appropriate matching. At the time of a scheduling decision, the COSHH algorithm uses the set of suggested job classes for each resource, and considers the priority, required minimum share,

and fair share of users to make a scheduling decision. Therefore, this algorithm takes into account heterogeneity in all three Hadoop factors that it has introduced.

Performance issues in BDCaM includes small job starvation, sticky slots, preserving fairness among users. Scheduling complexity also arise in such systems with different class of servers, arrival of jobs, class of jobs and number of servers.

Performance Metrics

There is a range of performance metrics that are of interest to both users and Hadoop providers. Five Hadoop performance metrics are used for evaluating the schedulers in this paper, including:

1. *Average Completion Time* is the average completion time of all completed jobs.
2. *Dissatisfaction* measures how much the scheduling algorithm is successful in satisfying the minimum share requirements of the users.
3. *Fairness* measures how fair a scheduling algorithm is in dividing the resources among users.
4. *Locality* is defined as the proportion of tasks which are running on the same resource as where their stored data are located.
5. *Scheduling Time* is the total time spent for scheduling all of the incoming jobs. This measures the overhead of each Hadoop scheduler.

B. Existing system

In the Existing System, they developed cloud-oriented context-aware middleware (CoCaMAAL). They described the context-aware service identification process using high level generalized medical rules. It proved its advantage for processing and managing large amount of contexts gathered from multiple AAL systems. It improved knowledge of understanding the patient's situation through iterative learning of present contexts and substantial historical data can reduce the transmission of repeated false alerts to the remote monitoring systems.

Limitations of existing system

It doesn't have (lacked) an important feature such as personalized knowledge discovery which could be derived from a large amount of patient data stored in the cloud repositories. It has lack of storage. It takes time to store the large amount of datas.

C. Proposed system

In this paper, we have proposed a BDCaM that is called Big data for Context-aware monitoring (Fig.1). This is an framework for individual healthcare monitoring process to predict the disabilities and abnormal conditions of particular patients. It has the advantages of context-aware computing, remote-monitoring, cloud computing, machine learn-ing and big data. It provides a systematic approach to support the fast-growing communities of people with chronic illness who live alone and require assisted care. The model also simplifies the tasks of healthcare professionals by not swamping them with false alerts. The system can accurately distinguish emergencies from normal conditions. It is mainly used for predicting the stronger relationship between vital signs and contextual information. It will make the generated data more consistent and the model will be more accurate for validation

Advantages of the proposed system

It has great accuracy during the prediction of correct abnormal conditions in a patient. It take less execution for predicting the abnormality by using context aware and daily activities of particular patient (Assisted living). It simplifies the tasks of healthcare professionals. The system can accurately distinguish emergencies from normal conditions.

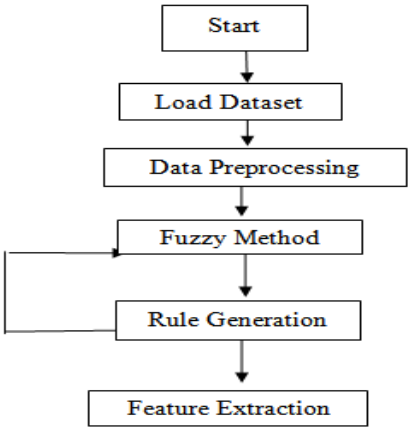

Fig 1:Proposed System

III. MODULES

The proposed system has the following five modules

1. Dataset Collection
2. Context Aggregator
3. Fuzzy Rule Extraction
4. Feature Extraction
5. Rule Matching and Classification

A. Dataset Collection

In this first we have to collect the dataset. Here we collect the dataset as ADL Dataset. ADL is nothing but an Activities of Daily Living (ADLs). This dataset comprises information regarding the ADLs performed by two users on a daily basis in their own homes. This dataset is composed by two instances of data, each one corresponding to a different user and summing up to 35 days of fully labeled data. Each instance of the dataset is described by three text files, namely: description, sensors events (features), activities of the daily living (labels). Sensor events were recorded using a wireless sensor network and data were labeled manually. We have to collect the dataset for different persons from different regions. Then we have to collect the profile information of users and activity logs of users and also we have to collect the medical records of particular persons.

Then, we have to preprocess the dataset. In this we have to eliminate the unwanted symbols or unwanted elements in the dataset.

B. Context Aggregator

After the dataset has been collected and preprocessed, data collector forwards the data to the context aggregator. In this we also collect the location and activity and environmental status of the particular person. Then we have to store the data's into the context aggregator cloud. The job of the context aggregator (CA) is to integrate all the primitive contexts in a single context state using a context model. Sometimes a single context attribute value as an individual has no meaning if it is not interrelated with other contexts. For example, an increment in HR seems an abnormal condition as a single context, but if the user doing exercise, this can be a normal situation. So, using past and present contexts, it can be determined whether the current user situation is normal or not. Therefore, all the contexts need to be aggregated to classify a situation accurately. The CA does this work and forwards the information to the context management system for the individual user.

C. Fuzzy Rule Extraction

An important feature of remote monitoring applications is to identify the abnormal conditions of a patient accurately and so send appropriate alerts to the care givers. In traditional systems, situations are classified by generalized medical rules or fuzzy rules which are not always applicable for every kind of patient. These systems cannot sense the future at an early stage. In some monitoring systems, when a patient feels unwell he/she needs to press a wearable panic button to notify a response centre about the emergency. A **fuzzy rule** is defined as a conditional statement in the form: IF x is A. THEN y is B. where x and y are linguistic variables; A and B are linguistic values determined by **fuzzy** sets on the universe of discourse X and Y, respectively.

D. Feature Extraction

A Context Management System (CMS) is the core component of the framework. The CMS consists of a number of distributed cloud servers that hold the big data. It stores the context histories of millions of patients. An important feature of remote monitoring applications is to identify the abnormal conditions of a patient accurately and so send appropriate alerts to the care givers.

E. Rule Matching and Classification

In the BDCaM model, the service providers are the cloud servers that sustain the generic medical rules to identify various types of diseases and symptoms. The rules of symptoms and anomalous behaviors are continuously updated by medical experts, doctors and other medical service providers. When any new rule is discovered in the CMS it also triggers the change in the SP cloud. The CMS uses rules of SP for data filtering and classification. When the CMS discovers any anomalous pattern in the context for a specific user it sends appropriate notification to the RMS. For example, when the BPlevel of a patient goes relatively high for a given situation, the CMS alerts the doctor to investigate it, but if it goes abnormally high then the CMS sends alerts to the emergency centre. Thus, the selection of RMS depends on situation classification. A major goal of our system is to classify a situation correctly to send proper alerts to the right RMS.

IV. CONCULSION

In this work, we have presented BDCaM, a generalized framework for personalized healthcare, which leverages the advantages of context-aware computing, remote-monitoring, cloud computing, machine learning and big data. Our solution provides a systematic approach to support the fast-growing communities of people with chronic illness who live alone and require assisted care. The model also simplifies the tasks of healthcare professionals by not swamping them with false alerts. The system can accurately distinguish emergencies from normal conditions. The data used to validate the model are obtained via artificial data generation based on data derived from real patients, preserving the correlation of a patient's vital signs with different activities and symptoms. The stronger relationship between vital signs and contextual information will make the generated data more consistent and the model will be more accurate for validation. The experimental evaluation of our system in cloud model for patients having different HR and BP levels has demonstrated that the system can predict correct abnormal conditions in a patient with great accuracy and within a short time when it is properly trained with large samples. In future, we intend to extend the model with more context domains.

REFERENCES

[1] A. Pantelopoulos and N. Bourbakis, "A survey on wearablesensor-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] D. N. Monekosso and P. Remagnino, "Behavior analysis for assisted living," IEEE Transactions on Automation Science andEngineering, vol. 7, no. 4, pp. 879–886, 2010.

[3] P. Groves, B. Kayyali, D. Knott, and S. Van Kuiken, "The bigdata revolution in healthcare," McKinsey & Company, 2013.

[4] S. Pandey, W. Voorsluys, S. Niu, A. Khandoker, and R. Buyya,"An autonomic cloud environment for hosting ecg data analysis services," Future Generation Computer Systems, vol. 28, no. 1,pp. 147–154, 2012

[5] A. Ibaida, D. Al-Shammary, and I. Khalil, "Cloud enabled fractal based ecg compression in wireless body sensor networks,"Future Generation Computer Systems, vol. 35, pp. 91–101, 2014.

[6] Australian bureau of statistics - 4821.0.55.001 - cardiovasculardisease in australia: A snapshot, 2004-05. [Online].

[7] [7] A. Forkan, I. Khalil, and Z. Tari, "Cocamaal: A cloud-orientedcontext-aware middleware in ambient assisted living," FutureGeneration Computer Systems, vol. 35, pp. 114–127, 2014.

[8] A. K. Dey, "Providing architectural support for buildingcontext-aware applications," Ph.D. dissertation, Georgia Institute of Technology, 2000.

[9] S.Sridevi, B. Sayantani, K.P.al Amutha, C. Madan Mohan, R.Pitchiah, "Context Aware Health Monitoring System, ICMB 2010, Springer LNCS 6165, pp. 249-257.

[10] S. B. Siewert. (2013, July) Big data in the cloud. [Online]. Available: http://www.ibm.com/developerworks/library/bdbigdatacloud/bd-bigdatacloud-pdf.pdf

[11] J. J. Oresko, Z. Jin, J. Cheng, S. Huang, Y. Sun, H. Duschl, andA. C. Cheng, "A wearable smartphone-based platform for realtime cardiovascular disease detection via lectrocardiogramprocessing," IEEE Transactions on Information Technology inBiomedicine, vol. 14, no. 3, pp. 734–740, 2010.