Mining Human Activity Patterns from Smarthome Big Data by using MapReduce Algorithm

[1]Dr. V. Karpagam [2]M. U. Pooja [3]M. Swathi
[1]Professor [2,3]Student
[1,2,3]Department of Information Technology Engineering
[1,2,3]Sri Ramakrishna Engineering College, Coimbatore, India

Abstract

Healthcare service is one of the most challenging aspects that is greatly affected by the migration of people to city centres. Cities are currently embracing massive digital transformation in an effort to support and provide a healthier environment. In such transformation millions of homes are being equipped with smart devices eg.smart meters, sensors etc which generate large volumes of data that can be analyzed to support health care services. The challenge is how to mine complex interdependencies among different appliances usage within a home where multiple data streams are occurring. The main goal is to discover human behavioural characteristics as an approach to understand and predict their activities that could indicate health issues. The human activity datasets which are generated by the smart meters are mined using the Big Data algorithms. If there is no usage of any appliances the result is given as an input to the health care application for further alerting needs.

Keyword- Hadoop on Big Data, MapReduce, Frequent Pattern

I. INTRODUCTION

As the world population is growing at a rapid pace, the demand for better health care facilities is increasing steadily. Earlier it was not always possible to monitor the patient continuously and for longer durations. A Healthcare service is one of the most challenging aspects that is greatly needed in many of the emergency situations which will be helpful for old age people and people who stay back at home.

Every day human metabolism changes according to their daily work or activities. User habits are mostly identified by everyday routines. Homes are being equipped with smart devices (e.g., smart meters, sensors, and so on) which generate massive volumes of numerical data such as current readings in electricity board that can be analyzed. Since all the habits are mostly identified by everyday routines, discovering these routines allow us to recognize anomalous activities. This system analyzes the energy consumption patterns using Hadoop MapReduce algorithm with the help of appliance usage data at smart homes, which is directly related to human activities.

A. Big Data

Big data is a blanket term for the non- traditional strategies and technologies needed to gather, organize, process, and gather insights from large datasets. While the problem of working with data that exceeds the computing power or storage of a single computer is not new, the pervasiveness, scale, and value of this type of computing has greatly expanded in recent years. An exact definition of big data is difficult to nail down because projects, vendors, practitioners, and business professionals use it quite differently. Generally speaking, big data is:

- Large datasets
- The category of computing strategies and technologies that are used to handle large datasets

Fig. 1: Big Data Analytics

In this context, "large dataset" means a dataset too large to reasonably process or store with traditional tool or on a single computer. This means that the common scale of big datasets is constantly shifting and may vary significantly from organization to organization.

B. MapReduce

MapReduce was designed by Google to efficiently carry out a set of functions against a large amount of data in batch mode. The "map" component distributes the programming problem or task across a large number of systems while managing placement to balance the load and allow recovery from failures. After the distributed computation is complete, another function called "reduce" aggregates all the elements back together to provide a result. An example of MapReduce would be determining the number of pages in a book that are written in 50 different languages.

C. Big Table

Big Table was developed by Google to be a distributed storage system to manage highly scalable structured data. Data is organized into tables with rows and columns. Unlike typical relational database models, Big Table is a sparse, distributed, persistent, multidimensional, sorted map. It has been designed to keep large volumes of data across commodity servers.

D. Hadoop Tool

Hadoop is an Apache-managed software framework created using MapReduce and Big Table. Hadoop allows applications based on MapReduce to run on large clusters of commodity hardware. The project has become the basis for the computing architecture underlying Yahoo!'s business. Hadoop is designed to parallelize data processing across computing nodes to speed computations and diminish latency. Two major components of Hadoop exist: a massively scalable distributed file system that can support petabytes of data, and a massively scalable MapReduce engine that computes results in batches.

II. RELATED WORK

Nowadays various data intensive applications have emerged, which serves the need for an efficient analytic model. The major benefits of MapReduce framework are scalability and fault-tolerance during massive data processing [1]. In terms of finding frequent patterns of human based on their daily activities, appliances data from smart current meters is used as the dataset where data mining algorithms are used to find the frequent patterns [3]. The combination of both BigData and smart systems can bring advanced healthcare systems within smart cities [2][4]. A pattern mining approach to sensor based human activity is also used for healthcare systems but the issue in such systems involves missing data because of sensor damage [5]. The proposed work uses both appliances data generated from smart data meter to find frequent pattern of human activities as dataset and Hadoop MapReduce algorithm to find any abnormal condition of the person who stays at home,with more accuracy and efficiency despite the volume of data. There is a training set which is used to generate frequent patterns from large appliance dataset, a testing set which is compared to the training set to find any abnormal activity of the person. In case of abnormal activity then it is assumed that the person possibly suffers from health problems, further raising alerts to inform the care providers or the neighbours.

III. PROPOSED SYSTEM

The proposed system aims to identify everyday human activities in smart homes and finding the frequent patterns. It analyzes energy consumption patterns of the appliances in smart home, which is directly related to human activities and discovers human behavioural characteristics as an approach to understand and predict their activities that could indicate health issues .The human activity datasets which are generated by the smart meters are mined using the Big Data algorithms. If there is no usage of any appliance, the result is given as an input to the health care application for further alerting needs.

A. Module Description

1) Extracting Frequent Patterns of Human Activities

Discovering human activity patterns from smart meter data. For example, activities such as Watching TV, Cooking, Using Computer, Preparing Food and Cleaning Dishes or Clothes are usually regular routines. The aim is to detect the patterns of these activities so that a health care application is able to monitor sudden changes in user's behaviour.

2) Clustering Analysis: Incremental k-means

Discovering the appliance-to-time associations is vital to health applications that monitor patients' activity patterns on a daily basis. In this section, a clustering analysis mechanism is used to discover appliance usage periodically. Appliance-to-time associations are underlying information in the smart meter time series data which include sufficiently close time- stamps, when relevant appliance has been recorded as active or operational. By using this data grouping a class or cluster of appliances that are in operation simultaneously or overlapping. The size of the cluster that describes such,

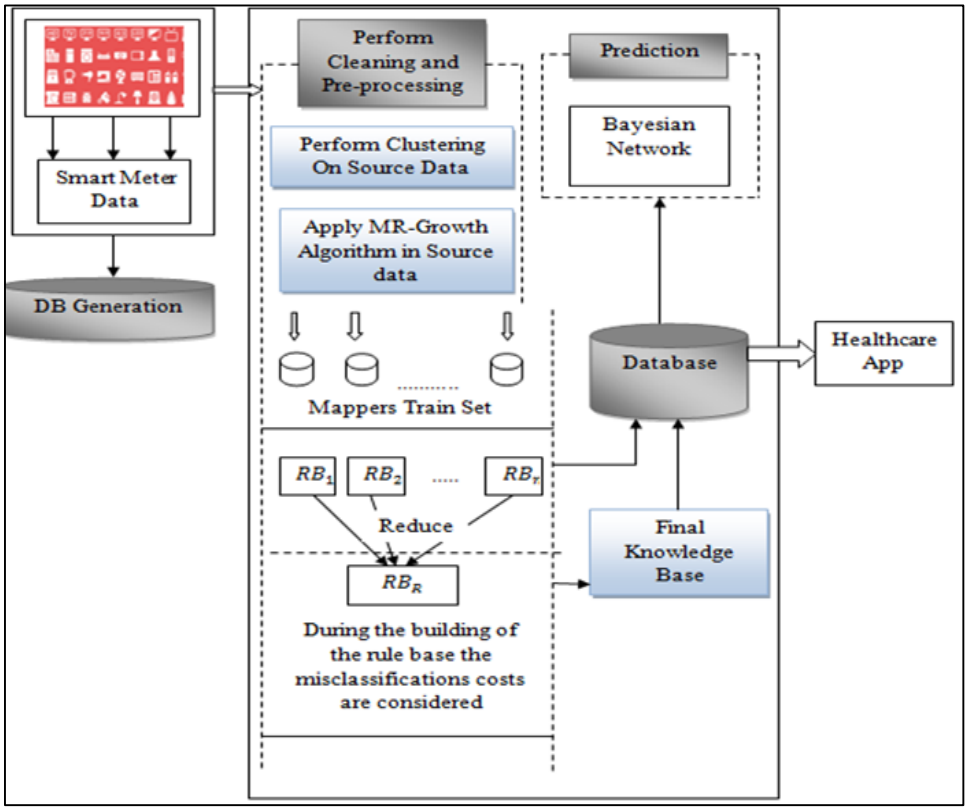

Fig. 1: System Architecture

The size of the cluster that describes such associations is defined as the count of members in the cluster as well as its relative strength. Clustering analysis is the process of creating classes (unsupervised classification) or groups/segments (automatic segmentation) or partitions where members must possess similarity with one another, but should be dissimilar from the members of the other clusters.

3) Activity Prediction Model

To integrate frequent patterns and appliance-to-time associations, to learn about the use of multiple appliances and build the activity prediction model. The mechanism utilizes Bayesian network which is a Directed Acyclic Graph (DAG), where nodes represent random variables and edges indicate probabilistic dependencies.

4) Map Reduce (MP) Growth Algorithm

The proposed MR-growth algorithm uses MapReduce to mine frequent patterns from huge amounts of uncertain data in a tree based pattern growth fashion.

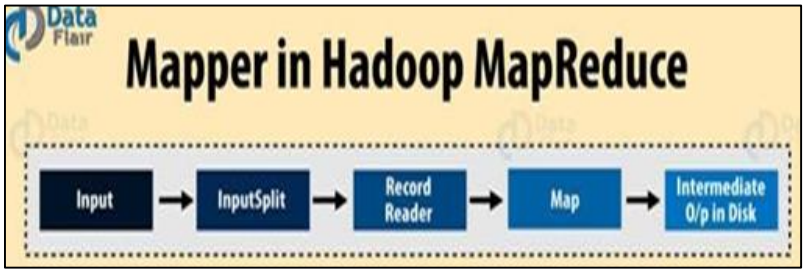

Fig. 2: Mapper

The algorithm can be divided into multiple stages. First, MR-growth reads a huge dataset of uncertain data. MR-growth divides the uncertain a set into several partitions and assigns them to different processes. During the Mapping phase, the mapper function receives each process, content of that process as input. For every process, the mapper function emits a key value <key,value> pair for each item. In mapper task, the output is the full collection of all these <key, value> pairs. No. of Mapper= {(total data size)/ (input split size)}

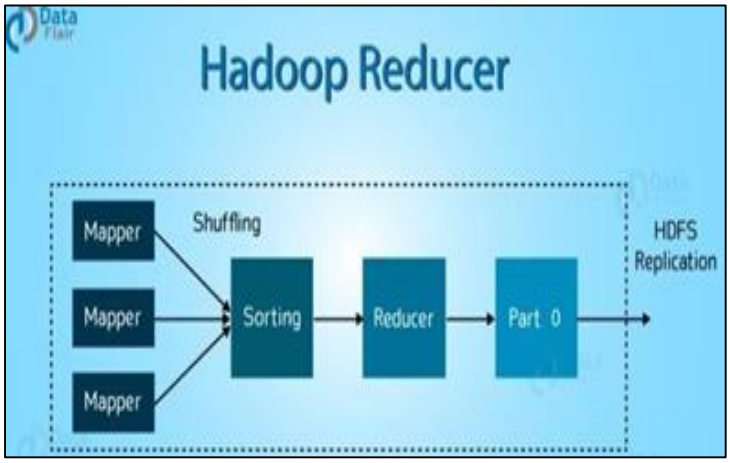

Fig. 3: Reducer

In reducer phase, Reducer takes the output of the Mapper (intermediate key-value pair) process each of them to generate the output. Hadoop Reducer takes a set of an intermediate key value pair produced by the mapper as the input and runs a Reducer function on each of them. One can aggregate, filter, and combine this data (key, value) in a number of ways for a wide range of processing. Reducer first processes the intermediate values for particular key generated by the map function and then generates the output zero or more key-value pair.

5) Performance Evaluation

It discovers, if there is no usage of any appliances or any abnormal activities by the user periodically and also find the similarity of appliances by using the Map-reduce growth algorithm.

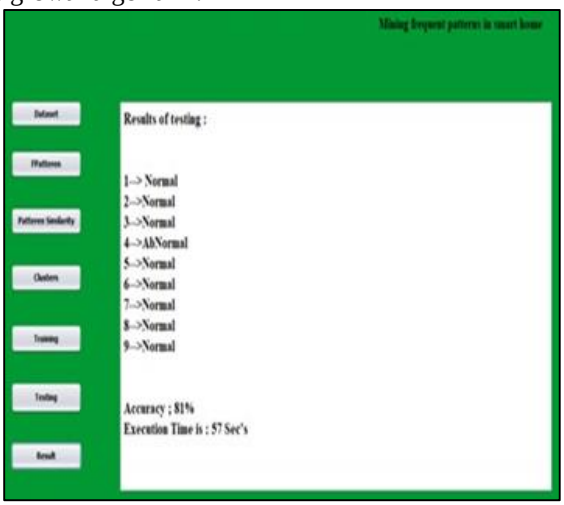

Fig. 4: Result page

The proposed model was implemented and tested using a testing dataset and the above result is obtained. The normal and abnormal state of the person is generated along with accuracy and execution time.

IV. CONCLUSION

This proposed idea on mining human activity pattern for healthcare for people who stay back at home or old age peoples using Big Data algorithm provide accurate results whether a person suffers with any health problem by mining his/her frequent activity patterns generated by the smart appliances.

REFERENCES

[1] Prasan kumar sahoo (senior member, ieee), Suvendu kumar "Analyzing Healthcare Big Data With Prediction for Future Health Condition"2017 ieee 7th International Advance Computing Conference.

[2] "Smart health: Big data enabled health paradigm within smart cities" by Pramanik, M. I., Lau, R. Y., Demirkan, H., & Azad, M. A. K.,2015 IEEE International Conference.

[3] "An Intelligent System for Mining Usage Patterns from Appliance Data in Smart Home Environment" by Yi-Cheng Chen, 2012 Conference on Technologies and Applications of Artificial Intelligence.
[4] "Estimating human interactions with electrical appliances for activity-based energy savings recommendations" by Cao, H. Â., Wijaya, T. K., Aberer, K., & Nunes, N 2016 IEEE International Conference on (pp. 1301- 1308). IEEE.
[5] "A pattern mining approach to sensor- based human activity recognition "by Gu, T., Wang, L., Wu, Z., Tao, X., & Lu, J. 2012 IEEE Transactions on Knowledge and Data Engineering, 23(9), 1359-1372