An Information Forwarder for Healthcare Service and analysis using Big Data

Aafaque Muzzam. A[1], Arunachalam. V[2], Awdhil Iththisham A.Z[3], Suresh. S

[1,2,3]Student, Dept. of Computer Science and Enginering, Panimalar Engineering College, Tamil Nadu, India
[4]Associate Professor, Dept. of Computer Science and Engineering, Panimalar Engineering College, Tamil Nadu, India

---

Abstract - *Wireless Body Area Networks (WBANs) are expected to play a major role in the field of patient-health monitoring in the near future, which gains tremendous attention amongst researchers in recent years. The advances in information technology have witnessed great progress on healthcare technologies in various domains nowadays. However, these new technologies have also made healthcare data not only much bigger but also much more difficult to handle and process. The data collected about the patients in remote healthcare applications constitutes to big data because it varies with respect to volume, velocity, variety, veracity, and value. The proposed scheme is based upon the initial cluster formation, retrieval, and processing of the big data in the cloud environment. Then, a fuzzy rule-based classifier is designed for efficient decision making about the data classification in the proposed scheme. The proposed scheme is evaluated on various evaluation metrics such as-average response time, accuracy, computation cost, classification time, and false positive ratio. Various clustering based algorithms and presented a set of recommendations regarding the use of various clustering techniques based on fuzzy c-means. These classifiers can be used for real-time remote healthcare applications which have not been exploited in the existing schemes. The results obtained are compared with the standard techniques to confirm the effectiveness of the proposed scheme. We also evaluate its performance in terms of energy consumption and communication/computation overhead.*

***Key Words*: Big Data Fuzzy rule- based classifier clustering algorithms average response time accuracy computation cost classification time**

1. INTRODUCTION

Big data in health informatics can be used to predict outcome of diseases and epidemics, improve treatment and quality of life, and prevent premature deaths and disease development. Big data also provide information about diseases and warning signs for treatment to be administered. This will help not only to prevent co-morbidities and mortality but also assists government to save the cost of medical treatment. It is very useful not only in clinical medicine for diagnosis/detection but also in epidemiological research as the big data will provide huge amount of data. The government, non-governmental organization and/ or pharmaceutical companies can use the data to formulate policies, strategies, intervention or medical treatment such as drugs development. Big data has implications on healthcare on patients, providers, researchers, health professionals.

Nowadays, there is an increasing demand for more information by the patients about their healthcare options or choices, and want participation in their health decision-making. The big data will help to provide patients with up-to-date information to assist them to make the best decision and to comply with the medical treatment.

Cloud computing is one of the fastest growing technologies of the modern era due to its usage in wide range of applications such as-remote healthcare, surveillance systems, weather forecasting, and efficient drug delivery systems.

All these applications generate a lot of data which needs to be handled carefully to have enhanced performance with respect to parameters such as-cost effectiveness, reliability, and availability.

In cloud computing environment, users have flexibility of resource access at any time as all the resources are located at a centralized location so that they can avail the functionalities of all these resources from anywhere. It allows the users to collect, access, and search the data with ease from anywhere even on the fly. In all of the above specified applications, data varies with respect to its size and types, as it is collected from different locations and devices. These devices may be located across the globe in a heterogeneous environment. The collected data varies in its velocity, volume, veracity, and value. So, the traditional storage and computation techniques may not be applicable to give a high level of performance with respect to parameters such as-response time, throughput, accuracy, and false positive ratio.

Moreover, the collected data from different devices is of spatio-temporal nature so, traditional database procedures to process such a large collection of data may not work well at all the time. Hence, instead of using the traditional storage and computing resources, there is a requirement of an efficient centralized repository such as cloud where all the data is stored and analyzed so that it can be accessible from anywhere at any time.

With an emergence of the Information and Communication Technology (ICT), remote diagnosis of the patients is now feasible. As the healthcare data varies in its size and volume, novel storage and retrieval methods should be used in healthcare related applications so as to provide a faster access of the data to the patients and doctors as per their requirements. Remote healthcare is gaining huge attention and with the evolution of ICT and cloud computing, this task can be realized practically. The data about the patients can be sent to the cloud from remote areas using ICT, where the information is extracted out of it and stored on various clouds based on its context.

This will help in better processing and fast retrieval of the results. Saria has analyzed that the amount spent on healthcare in US alone is about $3 trillion per year, but still it lags behind many countries in terms of quality of life and life expectancy. This cost could be reduced by the usage of efficient techniques used for big data analytics in healthcare applications in cloud environment. This big data can be analyzed using adaptive techniques and then processed data can be stored on the cloud and retrieved as per the demands of the users and doctors. As discussed above, the data collected in healthcare domain has properties such as-variety, velocity, and volume.

This is because the data is collected from multiple sources which can be static as well as mobile. Static sources consist of sensors that are generally placed inside the hospitals and homes from where the data is gathered and sent to the cloud. However, by using mobile sensors, a person can send the streaming data to the cloud even when that person is mobile by using body sensors. This streaming mobile data puts additional burden on the service provider as it needs to be processed in real-time. So, simple algorithms for handling such a large collection of data would not work well in this environment. Hence, there is a requirement of efficient mining algorithms to process the collected data from heterogeneous sources on the cloud servers.

2. SYSTEM ANALYSIS

2.1 EXISTING SYSTEM

- E-Health Applications in Internet of Vehicles
- Hidden Markov Model (HMM) to recognize the human behaviour using big data
- Clinical decision support system based on association rule mining
- Logistic regression model to predict dichotomous outcomes in healthcare data

2.1.1 DISADVANTAGE OF EXISTING SYSTEM

- Limited to only detecting cardiovascular diseases from compressed ECG data of patient and cannot be used to predict other diseases.
- Suffer from additional delay which is caused between transmitting the data and receiving the rules.

2.2 PROPOSED SYSTEM

- In Proposed system the framework is divided into three layers namely data acquisition layer, transmission layer, and computational layer.
- The cloud computing is used for management of health care services which is the basis of Healthcare as-a-service (HaaS).
- The cloud services are used for real time storage of data sensed from patients' body. This data is collected through various body sensors which are deployed in patient's body, or in the form of wearable sensors.

2.2.1 ADVANTAGE OF PROPOSED SYSTEM

- As the data is stored on cloud, simultaneous queries can be run on it to achieve high throughput with reduced delay.
- That doctor can input the symptoms of any new disease that breaks out and then finds only those persons who are likely to be get affected.
- A new entry is classified into clusters or when missing data value of the patient is collected.

3. REQUIREMENT SPECIFICATIONS

3.1 HARDWARE REQUIREMENTS

PROCESSOR	: Intel Core i3
RAM	: 4 GB DDR2 RAM
MONITOR	: 15" COLOR
HARD DISK	: 100 GB

3.2 SOFTWARE REQUIREMENTS

Front End	: ANDROID XML, JAVA
Back End	: SQLITE
Operating System	: Windows 07
IDE	: Eclipse, Android Studio

4. SYSTEM ARCHITECTURE

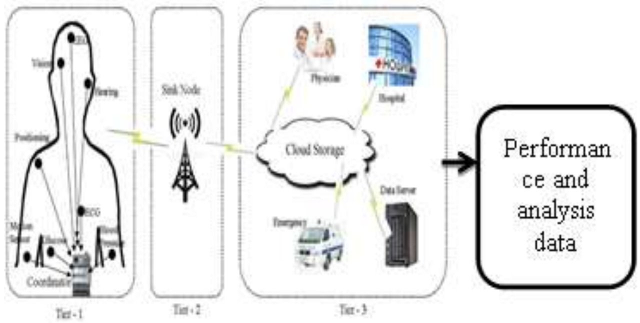

Figure 1 Architecture Diagram for Patient Healthcare System

5. PROJECT DESCRIPTION

5.1 GENERAL

This is a Requirements Specification document for a new web-based sales system for Solar Based Energy, Inc. (SBE). SBE is a distributor of alternative energy products including windmills, photovoltaics and fuel cells. The new system will upgrade the current websites to provide customers and employees customized browsing of the product catalog and the ability to complete product orders on-line. This document describes the scope, objectives and goal of the new system. In addition to describing non-functional requirements, this document models the functional requirements with use cases, interaction diagrams, and class models. This document is intended to direct the design and implementation of the target system in an object oriented language.

5.2 MODULE DESCRIPTION

5.2.1 DATA ACQUISITION

The data acquisition layer is responsible for gathering the data of the patients from different geographical domains. This data can be generated from the body sensor network, vehicular adhoc network, and network of devices present in homes or at various hospitals.The data from vehicular ad-hoc network is generated by the sensors placed in the vehicles such as hospital vans and ambulances and network connected to the cloud.

The data gathered from the networks (or devices) through transmission layer is stored at the cloud and results are computed in order to serve the requests from different clients. The cloud also contains the repository of data obtained from different hospitals.

5.2.2 CLUSTER FORMATION

The different clouds are divided into various clusters based on the gathered data. The doctor specifies the number of clusters to be formed at each cloudlet in the beginning and all the values that are provided to that cloudlet are classified into various clusters.

Once each cloudlet is divided into clusters and initial data values are stored in respective clusters, any new values (received at the cloudlet) are stored into clusters according to their membership. Whenever a new data value is sent to a cloudlet, its membership is checked with each cluster.

This new value is stored in a cluster which has maximum value of membership function with it. All the new values are stored in the same manner for each cloudlet

5.2.3 FUZZY CLUSTER FORMATION

Fuzzy cluster formation is used when the rules specified by the doctor are not crisp. For example, theBlood Pressure is specified as high. All such type of parameters are classified using fuzzy membership functions for initial clustering as well as for storing new data values of the patients. The fuzzification step is largely dependent on the shape of its membership function. Normally for faster computation, triangular membership functions are preferred.

The parameters such as alcohol, sp02level, blood pressure, and heart rate are fuzzy and the membership functions for only these input parameters are considered. These inputs are combined to form a fuzzy output using a specified linguistic rule-base. To calculate the value of fuzzy output, a defuzzification step is performed using the center of gravity (COG) method. It is the most commonly used method which gives accurate results for defuzzification.

5.2.4 RETRIEVAL OF RESULTS AND PERFORMANCE ANALYSIS

The doctor enters the query (symptoms of a disease) on the cloud. This query is divided into sub-queries which run on various cloudlets in parallel. Sub-query contains different rules. These rules are passed onto their respective cloudlets for processing whose results are processed. This significantly increases the processing speed because if the results are to be processed serially, then, one cloudlet has to wait for the results from another cloudlet.

The results of the sub-queries are combined in sequential manner in order to reduce the result set by just taking the common patients.In the end, the information in final result set is provided to the doctor which contains the list of probable patients for a particular disease.

6. MODULE EXPLANATION

The cloud is divided into a set of sub-clouds based on the parameters which are used to find out the list of patients who are suffering from a particular disease. For example, subcloud SA may be an age cloud, sub-cloud SB can be Blood Pressure (BP) cloud and similarly other sub-clouds can also be formed for other. These sub-clouds are the collection of various parameters that cover most of the diseases and these sub-clouds are further divided into different clusters.

The clustering is done based on the modified EM algorithm by calculating the membership of each value with specified clusters, and updating the clusters accordingly. For example, sub-cloud SA representing age cloud can be classified into three clusters of different age groups such as - infants, adults and old aged persons (Different clusters would be formed after consulting with an expert). Moreover, a fuzzy rule-based classifier has also been presented to classify fuzzy data and to retrieve fuzzy queries as specified by the expert. Once the clusters are formed, whenever a new record is received, cloudlets send different parameters to their respective sub-clouds, where data is stored in different clusters according to their membership value.

The detailed description about membership value of data is given in the next section. The major benefit of using this scheme of different sub-clouds for storing various parameters is that it classifies the records intelligently and retrieve the results quickly. For example, the doctor or the expert can specify the symptoms (in terms of parameters) and cloudlets will generate the results based on the parameters of the patients stored in various sub-clouds.

The novelty of our approach lies in the fact that whenever a new disease breaks out, the doctors can simply pass the symptoms for that disease in the system and the cloud will generate the probable patients which manifest those symptoms. A doctor can easily quantify the parameters according the symptoms of the disease. The parameters that cannot be exactly quantified are fuzzified and processed accordingly. The complete scheme works as follows. Initially, the different parameters are sensed from the patient as given.

These parameters are common and cover the symptoms of most of the diseases. It is to be noted that there can be additional parameters sensed from the patients; however, for the sake of simplicity only these parameters are considered in the proposed scheme.

The data about these parameters for individual patients is assumed to be gathered when the patient registers itself with the hospital. The doctor then inputs this data to his system (which can be updated from time to time) connected to the cloud. As this data is structured, different programmable context-aware cloudlets are assumed to be present at the cloud which segregates this data and sends it to respective sub-clouds. The quantifiable data gathered from patients gets accumulated in the clusters based on modified EM algorithm. The fuzzy classifier is used for retrieving the fuzzy rules specified by the expert. It is also used for storing fuzzy data values into clusters.

The inference engine used in the proposed scheme uses a rule-based searching for processing the results. The doctor passes the rules (as a query) to the inference engine according to symptoms of various diseases. The results are extracted from various clusters within the clouds and are sent back to the doctor. A unique patient id is associated with each result so that only those patient ids which fall in the premises specified by the doctor are presented in the results. However in extreme cases, it is possible that the information about a parameter (which is specified in the query by doctor) for a patient is not available at the cloud at a single point of time. In this case, this patient would not be included in the results of the query.

7. ALGORITHM EXPLANATION

7.1Cluster Formation:

Clustering is one of the routing design methodologies, used to effectively manage network energy and to apply aggregation techniques in the network the algorithms mostly evaluate the distance between a point and the cluster centroids to determine the cluster membership. The output from a clustering algorithm is basically a statistical description of the cluster centroids with the number of components in each cluster.

The cluster formation techniques have been extensively exploited in WSNs. Node eccentricity, energy level of nodes, weighted cost function, received signal strength, threshold time to form cluster, delay, number of iterations, stages of rounds, are the parameters considered in cluster formation. These were the main parameters considered in cluster formation by the previous researchers. It is employed with modifications, as per the application specific algorithms and protocols at present by the researchers.

7.2Storing of new values:

Once each cloud is divided into clusters and initial data values are stored in respective clusters, any new values which come at the cloud are stored into clusters according to algorithm initially, the individual membership function of the new value to each cluster is calculated. The maximum value of membership function is stored in a variable max and corresponding cluster number is stored in variable. The cluster having maximum membership function finally stores this new data value.

6. SAMPLE SNAPSHOTS

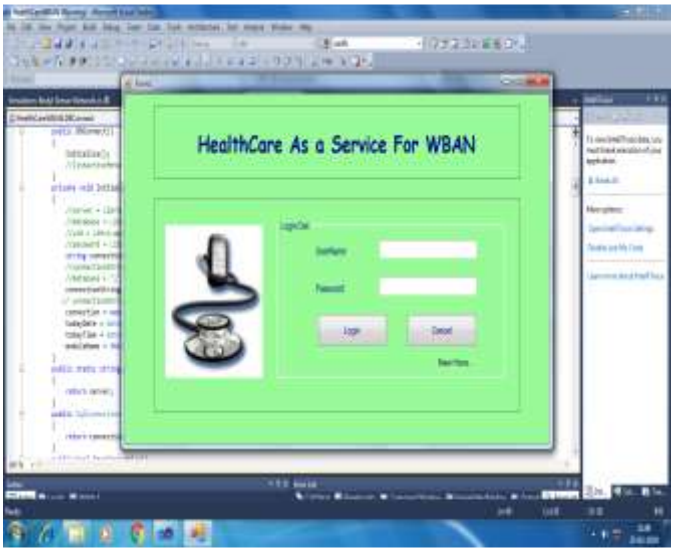

Figure 1 Login form of Admin

Figure 2 Login form of Patient

Figure 3 Recorded Inserted Success

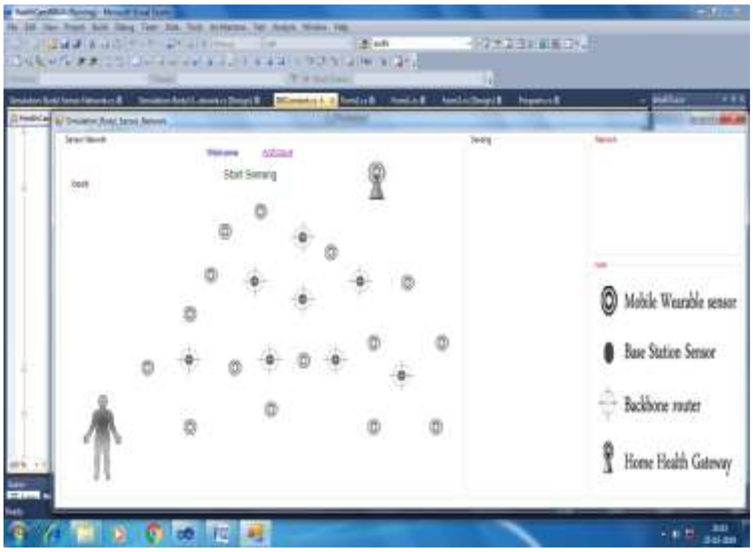

Figure 4 Sensing

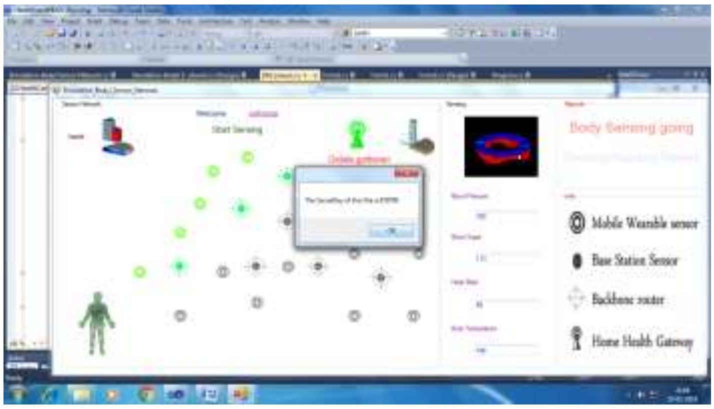

Figure 5 Security Key File

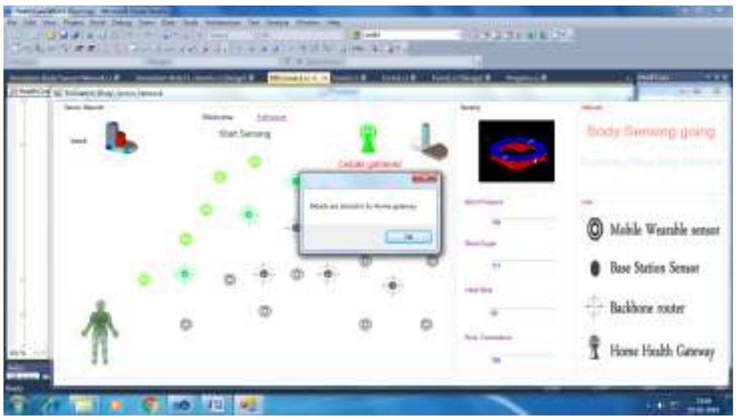

Figure 6 Home Gateway

7. CONCLUSIONS

As healthcare applications generate large amount of data which varies with respect to its volume, variety, velocity, veracity, and value, there is an imminent requirement of efficient mining techniques for context-aware retrieval and processing of this class of data. This paper proposes a new fuzzy rulebased classifier to provide Healthcare-as-a-Service (HaaS) and to classify the big data generated in this environment. The proposed scheme uses cloud-based infrastructure as a repository for storage and applying analytical algorithms for retrieval of information about the patients.

To apply analytics, algorithms for cluster formation and data retrieval are designed on the basis of Expectation-Maximization and fuzzy rule-based classifier. The proposed approach is compared with existing schemes and its performance is analyzed with respect to various evaluation metrics namely-average response time, accuracy, computation cost, classification time and false positive rate. The results obtained show that the proposed scheme is effective in finding out the probable patients suffering from a particular disease. Moreover, the proposed scheme performed better when compared with its counterparts namely multi-layer, Bayes network and decision table in terms of classification time and false positive rate.

8. REFERENCES

1. P. Jiang, J. Winkley, C. Zhao, R. Munnoch, G. Min, and L.T. Yang, "An Intelligent Information Forwarder for Healthcare Big Data Systems With Distributed Wearable Sensors," IEEE Systems Journal, vol. 10, no. 3, pp. 1147-1159, 2016.

2. Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, "Health- CPS: Healthcare Cyber-Physical System Assisted by Cloud and Big Data," IEEE Systems Journal, vol. 11, no. 1, pp. 88-95, 2017.

3. T.C. Havens, J.C. Bezdek, C. Leckie, L.O. Hall, and M. Palaniswami, "Fuzzy c-Means Algorithms for Very Large Data," IEEE Transactions on Fuzzy Systems, vol. 20, no. 6, pp. 1130-1146, 2012.

4. S. Fong, R. Wong, and A. Vasilakos, "Accelerated PSO Swarm Search Feature Selection for Data Stream Mining Big Data," IEEE Transactions on Service Computing, vol. 9, no. 1, pp. 33-45, 2016.

5. J. Stojanovic, D. Gligorijevic, V. Radosavljevic, N. Djuric, M. Grbovic, and Z. Obradovic, "Modeling Healthcare Quality via Compact Representations of Electronic Health Records," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 14, no. 3, pp. 545-554, 2017.

6. F. Sufi and I. Khalil, "Diagnosis of Cardiovascular Abnormalities From Compressed ECG: A Data Mining-Based Approach," IEEE Transactions on Information Technology in Biomedicine, vol. 15, no. 1, pp. 33-39, 2011.

7. N. Kumar, N. Chilamkurti, and J.H. Park,"ALCA: Agent Learning-based Clustering Algorithm in Vehicular Ad Hoc Networks," Personal and Ubiquitous Computing, vol.17, no. 8, pp. 1683-1692, 2013.

8. I.A.T. Hashem, I. Yaqoob, N.B. Anuar, S. Mokhtar, A. Gani, and S.U. Khan, "The Rise of "Big Data" on Cloud Computing: Review and Open Research Issues," Information Systems, vol. 47, pp. 98-115, 2015.

9. D. Lin, X. Wu, F. Labeau, and A. Vasilakos, "Internet of Vehicles for E-Health Applications in View of EMI on Medical Sensors," Journal of Sensors, 2015, DOI: 10.1155/2015/315948.

10. 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.