Cluster Based Wireless Mobile Healthcare System for Physiological Data Monitoring

Sagar B. Tambe\textsuperscript{a}, Ravindra C. Thool\textsuperscript{b}, Vijaya R. Thool\textsuperscript{a,b,*}

\textsuperscript{a}Department of Information Technology, SGGSIE&T, Nanded, Maharashtra, 431606, India
\textsuperscript{b}Department of Instrumentation Engineering, SGGSIE&T, Nanded, Maharashtra, 431606, India

Abstract

In recent years, importance in the medical healthcare system of Wireless Body Sensor Network (WBSN) has developed considerably. This paper discuss window based mining of vital signals, where the mining approaches are applied on the signal window in order to extract risk components and predict worst health condition. Also mining systems requires past prediction data of the patient in order to predict the future risk patterns, which is a bottleneck of the system where predicting the risk for new patient is impossible. This work also proves effective and efficient working of novel mining algorithm- ODRA (Online Distribution Resource Aware) used in healthcare.

Keywords: Wireless sensor network (WSN); Wireless body sensor network; Signal preprocessing; Patient monitoring; Mobile healthcare.

1. Introduction

The traditional patient monitoring systems are becoming less accurate as they work on the threshold values of the vital signals. For diagnosing many diseases with high accuracy, such simple threshold alarming system fails, because of the reason not all values which are above threshold for short duration of time are critical. The more detail study of patient monitoring says, if value of any physiological signal reaches to threshold (upper and lower bound given by the medical practitioners) and it remains in the same state for a longer duration, then it reflects the serious risk situation.

* Corresponding author. Tel.: +91-7588169570.
E-mail address: tambesagar@sggs.ac.in

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of organizing committee of the ICISP2015
Based on the same concept, the researchers are trying to build a system which would take a window of specific signals and monitor it by averaging the signal values within the window and comparing it with medical practitioner’s database. To achieve high accuracy in such windowing strategy there is a need to use data mining approaches. To predict the health risk by monitoring the physiological signal for a specific duration of time, the data mining approaches are necessary. The system requires grouping of the signals which shares some common attributes and then using robust prediction technique to predict the anomalies. Such system may need training data which are offline signals used to train the data mining system and generate rules, which can be used to accurately classify the new signals. For accurate diagnosis in all the times, patients need to be diagnosed in real time. To achieve this, the WBSN can be used which allows the real time patient monitoring no matter whether the patient is in the different ward, in different hospitals or even in different cities.

2. Real Time Historical Model

The historical model could give high accuracy for the patients whose data has been collected and mined, but it is not always possible to have historical data of the patient available in the system, may be due to patient visits first time. In such cases there is need of having new model which would work without referring to historical data and predict the risk with high accuracy. It works by monitoring the new vital signals from patients, applies real time stream mining approaches in order to predict the health risk. The crucial part of real time model is all computations are performed on mobile devices and notifications are given to physicians whenever immediate attention is required.

3. Real Time Signal Preprocessing

In signal processing data captured from sensors are expected to be dirty, a preprocessing phase has been implemented. In this processing signal do not address communication errors, because we assume to exploit protocols that grant data recovery and admissible time delay. Incompliant measurements are detected and removed by comparing them with two thresholds associated to the current signal and derived from medical literature. The effectiveness of the system is examined through the sensitivity analysis. It is found that a small number of clusters represented a good balance between within-cluster homogeneity and between-cluster heterogeneity.

4. Risk Components

The $z_1$ component used to measuring the health risk deriving from sharp changes in the physiological signal. It is obtained by considering the $h$-weighted slope over the time window $w$ offset by the most static condition.

$$z_1(x(t)) = \int_{t-w}^{t} h_{slope} | slope(x(t)) | \, dt - h_{\Delta} \Delta_w(x(t)).$$

The most static condition for a signal in the time window is defined as the smoothest change between its first value $x(t-w)$ and its current value $x(t)$

$$\Delta_w(x(t)) = |x(t) - x(t-w)|$$

The $z_2$ component measures the risk deriving from the $h$-weighted offset over the time window. While $z_1$ focuses on quick changes, $z_2$ evaluates long-term trends, as it is offset-based

$$z_2(x(t)) = \int_{t-w}^{t} h_{offset} | offset_w(x(t)) | \, dt.$$
5. Real Time Clustering

The ODAC (Online Divisive-Agglomerative Clustering) system is a clustering algorithm that designs a tree-shaped structure of clusters using a upper-down strategy. The system encloses an incremental instance measure and split for expansion and aggregation of the hierarchical structure, based on clusters diameters. If a given heuristic condition is met on this diameter of cluster, the system splits the cluster and assigns each of the chosen variables to one of the new clusters, becoming this as the pivot variable for that cluster. The algorithm uses partition based approach and inspired from Resource Aware High Quality Clustering (RAH Clustering). After formation of new cluster, the closest pivot assigned to new cluster which is taken from old cluster. New node start new statistics, considering that only upcoming information will be useful to decide splitting of data for new cluster. Also validate the approach through real system implementation with help of real sensor nodes such as heart rate (HR) sensor and pulse oximeter (Spo2) for physiological data. This feature increases the system's ability to scope with new concepts as test is performed such that if the diameters of the sub node leaves approach the parent's diameter.

6. The partition based clustering algorithm K-means

K-means algorithm is simplest learning algorithms that used to solve all clustering problem. In this K-means algorithm to define k centroids for single cluster. Then cluster take each point belonging to a given data set and associate it to the close centroid. Then re-calculate k new centroids as centers of the clusters resulting from the starting point. After we have these k new centroids, a new binding has to be done between the same data set points and pass to new centroid of algorithm. After result we notice that the k centroids change their position step by step until no more changes are done. In this clustering algorithm large streaming data, overload data mining systems, which can lead to less performance and intolerable processing delay for real-time applications. Finally, this algorithm aims at reducing an objective function, in this case a squared error function. The objective function

\[ J = \sum_{j=1}^{k} \sum_{i=1}^{x} \| x_i^{(j)} - c_j \|^2 \]  

Where, \( \| x_i^{(j)} - c_j \|^2 \) is a chosen distance measure between a data point \( x_i^{(j)} \) and the cluster centre \( c_j \) is an indicator of the distance of the \( n \) data points from their respective cluster centres.

7. Functions

7.1 Validity checks on input

The inputs to the system are vital signal values. These input values are pre-processed by applying proper validity check on them. If the value of the signal is noisy, it is either skipped or predicted from its past average.

7.2 Exact sequence of operation

The system shall be designed by robust state transition mechanism so if any operation fails, or system violates the sequence of operation it must restart without having the past prediction dropped out.
8. Architecture of the system

The real time model is contribute that existing historical data, where the risk level is predicted with or without knowing the past data records of the patients and the critical alerts are informed to the physician in case of emergency. To achieve high accuracy in real time, we propose innovative stream mining algorithm. Each of the roles represents a person that uses a smart device such as a mobile phone to communicate with the server setup in the care centre such that he or she can go around without restrictions.

9. Security Issues and Privacy of system

The security issues in a WBSN differ in several aspects from those in other applications. In medical applications there are some typical issue is to identify an attack from network failure or network congestions. Therefore detection technologies are required. Guided by the OSI model layering-based attacks on the protocol stack may be countered by the combined security model. A various type of wireless body sensors connect to an embedded control system security related problems are much simpler. Security and Privacy both approaches convey with an android smartphone via Bluetooth, ZigBee or Wi-Fi module. Because of the all over the place use of Bluetooth or Wi-Fi in smart phone communication various hazard or threat exist. Threat reduction involves accurately designed conversion between the sensor network and the smartphone, that is a defined process of software engineering, authentication, verification and use of the encryption, comprehensive testing. Nowadays, android is one of the most famous operating system for mobile phones. The android platform stack is created on a Linux kernel system. The android security framework is based on the Linux security system, android specific security technology and the cellular network provider that is application permissions, signing applications and component encapsulation.
10. Online Distribution Resource Aware (ODRA) Clustering

The ODRA algorithm is efficient, resource aware and accurate algorithm which is specially designed for Ubiquitous computing especially for mobile platforms. ODRA eliminates adding of new clusters from RAH Clustering, hence put limit on number of clusters that are formed. This is because system requires maximum three numbers of clusters for the three risk levels. Following notations are used for structures related to cluster center.

\[ c(j) \] – (where 1 <= j <= k) Cluster center j,
\[ c'(j) \] – Cluster j for all sum of vector point,
\[ q(j) \] – Cluster j for all assigned number of points,
\[ p(j) \] – Last moved distance that c(j),
\[ s(j) \] – Its closest distance center from c(j).
\[ x(i) \] – (where 1 <= i <= n) data point i.
\[ a(i) \] – x(i) is assigned index of the center.
\[ u(i) \] – Its assigned center c(a(i)) and upper bound on the distance between x(i).
\[ l(i) \] – Its second closest center to x(i) that is not c(a(i)) and Lower bound on the distance between x(i) and
\[ u(i) \geq |x(i) - c(a(i))|\]
\[ l(i) \leq \min_{j \neq a(i)} |x(i) - c(j)| \]

1. CALL INITIALIZE
2. While not converged do
3. For each c_j do
4. Update s_j by second closest cluster centers
5. For all data points x_i
6. If upper bound u(i) is greater than lower bound l(i) and T then
7. Update u(i) by distance between x_i and center c_j to which x_i is assigned c(a(i))
8. Assign a' as index of center to which x_i is assigned a(i)
9. CALL POINT-TO-ALL-CTRS
10. If a' \neq a(i) and distance between data point x_j and its center c_j is less than T then, update
11. Else if distance between data point x_i and its center c_j is greater than T then
12. Assign m' = x_i
13. Compute U_m;
14. Assign R_m as ratio of difference between total memory size Nm and memory usage Um and Nm
15. If R_m < \frac{1}{LB_m} then
16. Decrement T;
17. Update W;
18. If (1 - R_m) < 20\% then
19. Increment T;
20. Update W;
21. CALL MOVE-CENTERS
22. CALL UPDATE-BOUNDS

11. Mode of System Operation

To measure performance of the system the system works in three modes of the operation and for both RAH and ODRA clustering algorithms.

11.1 Historical (HT)

In this mode of operation, system does not execute the real time clustering; instead it maps the risk components z_1, z_2 and z_3 to historical rule base and predicts the risk level. The accuracy of the HT depends upon
the presence of patient or disease specific data of the particular patient being monitored or HT gives high accuracy if the patients have their past record in the database.

11.2 Real Time (RT)

WSN have the potential to revolutionize the capture, communication, and processing of critical data for use by first user\textsuperscript{12}. In this mode of operation system apply online clustering algorithms on the risk components, and in order to form three clusters where each cluster centre will be mapped to real time risk function to get the risk level. The system is operated in this mode for both clustering algorithms and their performance in terms of running time, memory requirement and SSQ (Sum of Squared Distance)

12. Performance Analysis of Clustering Algorithms

The performance analysis of the clustering algorithm is evaluated by measuring running time, memory requirement and average SSQ. The analysis is applied on batch of different data points and for varying window size.

12.1. Running Time

Table 1. Time Complexity of Real Time Clustering Algorithms for Window Size =10

<table>
<thead>
<tr>
<th>No. of data points processed</th>
<th>RA-HCluster (in milliseconds)</th>
<th>ODRA (in milliseconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>312</td>
<td>234</td>
</tr>
<tr>
<td>200</td>
<td>624</td>
<td>655</td>
</tr>
<tr>
<td>250</td>
<td>780</td>
<td>140</td>
</tr>
<tr>
<td>260</td>
<td>811</td>
<td>187</td>
</tr>
<tr>
<td>265</td>
<td>811</td>
<td>202</td>
</tr>
<tr>
<td>270</td>
<td>842</td>
<td>842</td>
</tr>
<tr>
<td>300</td>
<td>936</td>
<td>936</td>
</tr>
<tr>
<td>400</td>
<td>248</td>
<td>248</td>
</tr>
<tr>
<td>500</td>
<td>560</td>
<td>560</td>
</tr>
</tbody>
</table>

Fig.2. Time Complexity Graph for Window Size = 10
12.2 Average SSQ

Table 2. Average SSQ of Clustering Algorithm for Window Size = 10

<table>
<thead>
<tr>
<th>No. of data points processed</th>
<th>RA-H Cluster</th>
<th>ODRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>1.426303</td>
<td>1.0367</td>
</tr>
<tr>
<td>200</td>
<td>3.271483</td>
<td>1.550124</td>
</tr>
<tr>
<td>250</td>
<td>4.678134</td>
<td>3.568425</td>
</tr>
<tr>
<td>260</td>
<td>2.833679</td>
<td>1.965847</td>
</tr>
<tr>
<td>265</td>
<td>2.872582</td>
<td>2.458746</td>
</tr>
<tr>
<td>270</td>
<td>1.461093</td>
<td>1.263587</td>
</tr>
<tr>
<td>300</td>
<td>3.046561</td>
<td>2.547852</td>
</tr>
<tr>
<td>400</td>
<td>5.275355</td>
<td>5.048596</td>
</tr>
<tr>
<td>500</td>
<td>2.695268</td>
<td>1.958647</td>
</tr>
</tbody>
</table>

Fig. 3. Average SSQ Graph for Window Size = 10

12.3 Memory Utilization

Table 3. Memory Utilization of Clustering Algorithm for Window Size = 10

<table>
<thead>
<tr>
<th>No. of data points processed</th>
<th>RAH Cluster Memory (Kb)</th>
<th>ODRA Memory (Kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>3217</td>
<td>82</td>
</tr>
<tr>
<td>200</td>
<td>4502</td>
<td>2698</td>
</tr>
<tr>
<td>250</td>
<td>485</td>
<td>410</td>
</tr>
<tr>
<td>260</td>
<td>6294</td>
<td>5794</td>
</tr>
<tr>
<td>265</td>
<td>3342</td>
<td>5813</td>
</tr>
<tr>
<td>270</td>
<td>3767</td>
<td>1667</td>
</tr>
<tr>
<td>300</td>
<td>3448</td>
<td>5477</td>
</tr>
<tr>
<td>400</td>
<td>3770</td>
<td>4985</td>
</tr>
<tr>
<td>500</td>
<td>7967</td>
<td>8265</td>
</tr>
</tbody>
</table>
Conclusion and Future works

In system operation, signal samples from different age groups and diseases have applied in historical, real time and historical with real time mode and for HR signals it is observed that, historical mode of operation gives average accuracy near to 91% where the samples are pre-mined in the system, whereas in real time mode, system gives accuracy near to 94% for pre-mined as well as new samples and the historical with real time mode leads to highest average accuracy 96%. The historical with real time mode of operation implemented with ODRA gives highest accuracy 96% than RAH clustering where the accuracy is 94%. It can be proved that working of historical with real time mode is efficient and effective than other modes, which is the need of risk prediction system; hence proposed system is best suited in healthcare domain where accuracy is major concern. In future the risk levels alerts can be sent to the expert diagnosis system, by which the primary diagnosis or first aid diagnosis can be informed to the patients, where the physician may take long time to reach for diagnosis.

References