Discovering Vital Trends For Personalized Healthcare Delivery

Rahul Krishnan Pathinarupothi

Amrita Center for Wireless
Networks and Applications,
Amrita University, Kollam, India
rahulkrishnan@am.amrita.edu

Ekanath Rangan

Amrita School of Medicine,
Amrita University, Cochin, India
ammasekanath@gmail.com

Abstract

Personalization of remote health monitoring and healthcare delivery is a challenging research problem faced by practitioners and researchers alike. In this paper we present techniques for trend analysis and data summarization using personalized severity based motif discovery in large time series medical data.

Author Keywords

Motif discovery; remote health monitoring

ACM Classification Keywords

• Information systems~Decision support systems
• Information systems~Data analytics

Introduction

Use of wearable sensors and IoT devices for continuous monitoring of body vitals is generating great interest in the research community. However, we are yet to see large-scale adoption of such devices and techniques in use for day-to-day remote monitoring of patients. One of the major challenges in this area of research is the development of effective tools and algorithms that can summarize the voluminous sensor data and present it in a meaningful way to the practitioners. Many a time the signals obtained from body worn sensors need to be analyzed according to previous trends of the respective patients and not just deviations from

Need for Personalization

For patients having normal blood pressure (BP) levels, it might require critical support if the BP measurement goes above high thresholds, whereas for patients with history of hypertension, this threshold might be even higher. Similarly, for other vital sensors too, there is a need for personalizing data summarization using patient specific trends and thresholds with active feedback from the doctors.

medically accepted thresholds. Due to these intra-patient variations in identifying abnormal vitals measurement, it becomes important to personalize the summarization of sensor data too. In this paper, we argue for using computationally simple algorithms and techniques for smart summarization of large medical time-series data. We have developedtechniques and algorithms to identify normal and disease patterns that we call as *Consensus Motifs* (CM) and *Consensus Abnormality Motifs* (CAMs) respectively. We call this frameworkamespace**RASPRO**, which stands for **RApid Summarization for effective PROgnosis**.

Sensing architecture

We consider N vital sensors connected to a patient's body. The data is first passed through the RASPRO quantizer (see Figure 1) that converts the continuous values from the sensors to K different discrete severity level symbols based on medically accepted severity bands. The normal values are assigned the symbol A, while above-normal and subnormal values are quantized into A++, A+, A-, A-- etc. respectively; + symbol representing the above normal and - symbol representing subnormal values. The doctors based on his/her assessment of the patient as well as on symptomatic correlation with the sensor values can alter these levels. The RASPRO quantizer passes the quantized time series data to the multiplexer (MUX), where these are arranged into a single sensor matrix (SSM). In this, the quantized values are arranged as increasing time series data in f columns and w rows, where f = frequency of sensing and w = observation window.Now, we use the SSM to derive two different summaries of the entire w x f data points; first the most frequent normal vital data (CM), and second the most frequent abnormality data (CAM).
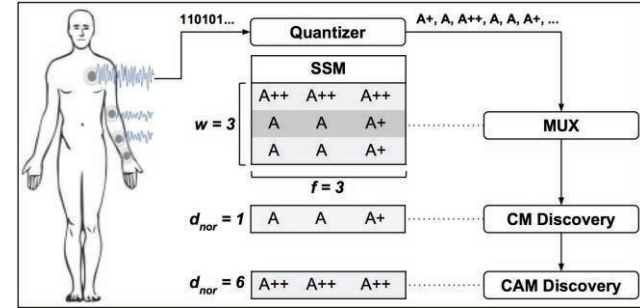

Figure 1: RASPRO architecture showing consensus motif (CM) and consensus abnormality motif (CAM) detection.

Consensus Motif Discovery

By analyzing how doctors analyze data on a day-to-day basis, it was observed that they are more interested in sequences of data that have strong temporal connection and regular intervals between them. To make it further clear, in many cases the doctors analyze the blood pressure (BP) data across a seven day period by comparing the morning BP data with other morning BP data and so on. This analysis pattern makes consensus motif discovery a much less computationally complex problem. So, the first row of sequences in SSM is compared with the other rows in an ordered row wise comparison to find the most frequent data trend. We now describe a deterministic brute force method for discovering CM from an SSM.

Let us call each row of the SSM as a candidate motif (M_{CAN}).First, for each of these M_{CAN}s, the sum of hamming distances Σd_h from every other M_{CAN}in the SSM is calculated. The M_{CAN} with the minimumsumis selected as the consensus motif, since it represents the centroid of the cluster of all the data points. The CM thus obtained (refer **Table 1**) represents the overall trend of the vitals for the time window w.

Motif Discovery

There have been many previous works [2], [3], [4], [5], [6] on motif discovery in time series data, which are both time consuming as well as computationally expensive. However, we define the problem of motif discovery in medical sensor data in a much more lucid terms based on the analytical pattern useful for the practitioners.

Consensus motif represents the centroid of the most frequently occurring M_{can}, and hence it shows the vitals trend. The discovery of a consensus motif forms the basis for summarization of data.

Multiple CMs

We might obtain multiple CMs after the discovery process, though the chances of this are very less since the vitals signs usually express some common trend.

sBP SSM (f = 3, w = 7)					
Day	T1	T2	T3	Sum (d_h)	CM
1	A	A	A+	5	✔
2	A	A	A+	5	✔
3	A+	A	A+	7	✗
4	A	A+	A	12	✗
5	A	A	A	9	✗
6	A+	A	A+	7	✗
7	A+	A	A+	7	✗

Table 1: Systolic BP (sBP) of a patient for 3 days and the corresponding calculation of hamming distances and CMs.

To ensure that we do not consider frequently occurring high severity data, we consider M_{CAN} that are only 2 hammingdistanceor less away from the medically accepted normal threshold (represented by the motif <A, A, A>).The last column in Table 1 lists if a given M_{CAN} can be a representative CM for the entire SSM. For this data, the motif <A, A, A+> is considered as the CM.We also note that the brute force consensus motif discovery method involves $O(w^2)$ comparisons.

We now focus on an equally important task of finding abnormalities in the vitals data that is extremely significant in prognosis. We define a term called *Consensus Abnormality Motif (CAM)* to represent this, and it is defined as the most frequently occurring abnormality pattern in a given SSM. Next, we describe a brute force method to discover CAM.

CAM discovery

The brute force method starts with the calculation of hamming distance for each candidate motif M^1_{CAN}, M^2_{CAN},

...,$M^w_{CAN}$from the normal motif <A, A, A>. If a M_{CAN}has hamming distance greater than a given medically accepted threshold, say d_{nor}, then we proceed to calculate the sum of hamming distances Σd_h of that M^n_{CAN} from all other candidate motifs. At the end of this analysis, we will have one or more CAMs whose Σd_h is the minimum. The CAMs with minimum represent the most common abnormalities, while the CAMs with highest hamming distance from <A, A, A>show the most severe conditions.One of the disadvantages with this method is that multiple CAMs, for instance <A+, A+, A+> and <A+++, A, A> both are three hamming distance away from<A, A, A>. However, the severity of A+++ is much higher compared to A+. In many cases, such severe situations might be missed out as sub-severe conditions using the above said method. To overcome this, we introduce a severity weighted hamming distance (wd_{nor}), where each severity level has an associated weight. For instance, weights for A+, A++ and A+++ are assigned as 1, 2 and 3 respectively. So, the hamming distance of <A+, A+, A+> from M_{nor} is 3, while that of <A+++, A, A> is 9.

sBP SSM (f = 3, w = 7)							
Day	T1	T2	T3	Sum (d_h)	d_{nor}	CAM	wd_{nor}
1	A+	A+	A+	15	3	✗	3
2	A+	A+	A	11	2	✔	2
3	A+++	A	A	13	3	✗	9
4	A+++	A	A	13	3	✗	9
5	A+	A+	A	11	2	✔	2
6	A+++	A	A	13	3	✗	9
7	A+	A+	A+	15	3	✗	3

Table 2: Representing the utility of using weighted hamming distance for discovering severe vital trends.

CAM Discovery

As seen in the CM discovery, the problem of CAM discovery is also simplified using SSM and in effect we are comparing each row of SSM with other rows.

Complexity

The brute force CAM discovery method involves $O(w^2)$ comparisons too.

Table 2 shows an interesting data set where the CAM is <A+, A+, A> whereas the most serious occurrences of M_{can}<A+++, A, A> (marked in red) hides along with <A+, A+, A+> (marked in blue) when using d_{nor} measurement alone. Using the weighted hamming distance approach, we see that the serious M_{can}s can be differentiated from less severe signals.

Conclusion and Future Work

In this paper we have presented an initial work on a the development of a simple, yet effective data summarization technique that can help in personalizing remote healthcare delivery systems according to medically accepted severity thresholds as well as patient specific characteristics. We call this framework as RASPRO. Currently, we are readying for large-scale testing of the algorithms and techniques using patient data from one of the largest hospital information systems, Amrita HIS [1], deployed at the Amrita Institute of Medical Sciences and Research Center in Kochi, India.

Acknowledgements

References

1. Achan, P.P., Healthcare Charities, Inc., 2011. System and method to develop health-care information systems. U.S. Patent 8,000,977.

2. Balasubramanian, A., Wang, J. and Prabhakaran, B. Discovering Multidimensional Motifs in Physiological Signals for Personalized Healthcare. Selected Topics in Signal Processing, IEEE Journal of, (2016).

3. Chiu, B., Keogh, E. and Lonardi, S., 2003, August. Probabilistic discovery of time series motifs. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 493-498). ACM.

4. Hajihashemi, Z. and Popescu, M., 2015, August. Early illness recognition using frequent motif discovery. In 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 3699-3702). IEEE.

5. Liu, Bo, Jianqiang Li, Cheng Chen, Wei Tan, Qiang Chen, and MengChu Zhou. Efficient Motif Discovery for Large-Scale Time Series in Healthcare. Industrial Informatics, IEEE Transactions on. (2015): 583-590.

6. Lonardi, J.L.E.K.S. and Patel, P., 2002. Finding motifs in time series. In Proc. of the 2nd Workshop on Temporal Data Mining (pp. 53-68).