Computer Networks 113 (2017) 244–257

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

ViSiBiD: A learning model for early discovery and real-time prediction of severe clinical events using vital signs as big data

Abdur Rahim Mohammad Forkan [a,b,*], Ibrahim Khalil [a,b], Mohammed Atiquzzaman [c]

[a] School of Science (Computer Science), RMIT University, Melbourne, Victoria 3001, Australia
[b] National ICT Australia (NICTA), Australia
[c] School of Computer Science, University of Oklahoma, Norman, OK 73019-6151, United States

ARTICLE INFO

Article history:
Received 21 April 2016
Revised 24 December 2016
Accepted 30 December 2016
Available online 3 January 2017

Keywords:
Big data
Vital sign
Cloud computing
Correlations
Knowledge discovery
Data mining

ABSTRACT

The advance in wearable and wireless sensors technology have made it possible to monitor multiple vital signs (e.g. heart rate, blood pressure) of a patient anytime, anywhere. Vital signs are an essential part of daily monitoring and disease prevention. When multiple vital sign data from many patients are accumulated for a long period they evolve into big data. The objective of this study is to build a prognostic model, **ViSiBiD**, that can accurately identify dangerous clinical events of a home-monitoring patient in advance using knowledge learned from the patterns of multiple vital signs from a large number of similar patients. We developed an innovative technique that amalgamates existing data mining methods with smartly extracted features from vital sign correlations, and demonstrated its effectiveness on cloud platforms through comparative evaluations that showed its potential to become a new tool for predictive healthcare. Four clinical events are identified from 4893 patient records in publicly available databases where six bio-signals deviate from normality and different features are extracted prior to 1–2 h from 10 to 30 min observed data of those events. Known data mining algorithms along with some MapReduce implementations have been used for learning on a cloud platform. The best accuracy (95.85%) was obtained through a Random Forest classifier using all features. The encouraging learning performance using hybrid feature space proves the existence of discriminatory patterns in vital sign big data can identify severe clinical danger well ahead of time.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

The connected world is creating mountains of data that are populated in the cloud storage [1] and analysed using the high processing power of cloud computing technology [2] to discover useful patterns for future behaviour prediction. The adoption of body and wireless sensor network technology [3,4] along with the emergence of pervasive computing applications [5] in home-based health monitoring [6,7] has dramatically increased the quantity of medical data that are available electronically [8] via cloud platforms. Vital signs such as heart rate (HR), blood pressure (BP), respiratory rate (RR), oxygen saturation (SPO_2), body temperature, and ECGs are a crucial part of these medical data [9]. For example, if we consider the numerical value of each vital contains 4 bytes and the frequency of data collection is 1 minute, then for 6 vital signs total 24 bytes data gathered per minute, which is equivalent to 33.75 KB per day, or 12 MB per year[1]. If such data are gathered from 5 million patients, then the data amount will be 57.3 PB per year[2]. This statistics is only for 6 vital signs data. Along with these numerical vital signs if we consider to collect data of other bio-signals such as ECG, PPG, EEG which are high in resolution then data amount will be several exabytes per year because a 12-lead ECG device alone generates 1.98 GB [10] and 64-channel EEG device generates 2.1 GB [11] per day and consequently such data will rapidly reach several zetabytes.

Vital signs can be described by the primary characteristics of big data [12] as shown in Fig. 1. Vital signs from varieties of patients are accumulated continuously, resulting in an incredible *volume* of data. Cloud computing facilitates more effective capture, storage and manipulation of this large volume of data. Since vital sign data are collected in real-time and at a rapid pace (i.e. samples per minute) they have *velocity*. Vital signs and their correla-

* Corresponding Author.
E-mail addresses: abdur.forkan@rmit.edu.au, s3361322@student.rmit.edu.au (A.R.M. Forkan), ibrahim.khalil@rmit.edu.au (I. Khalil), atiq@ou.edu (M. Atiquzzaman).

[1] $24 \times 60 \times \times 24 \times 365 = 12$ Megabytes.
[2] $12.03 \times 5 \times 10^6 = 57.36$ Petabytes.

http://dx.doi.org/10.1016/j.comnet.2016.12.019
1389-1286/© 2017 Elsevier B.V. All rights reserved.

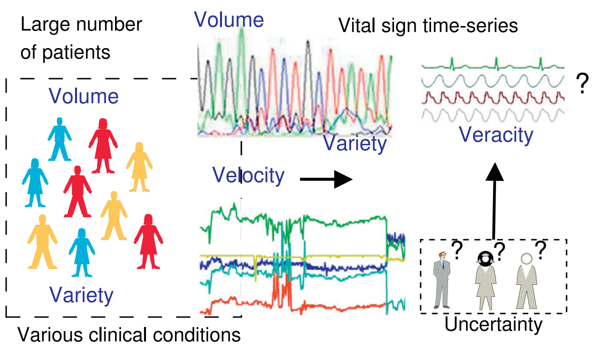

Fig. 1. Vital signs as big data.

tions are also different in nature in different patients and so they have *variety*. Moreover, there is a high degree of uncertainty in correlations of these evolving data streams which mean they have *veracity*.

The changes in multiple vital signs indicate several symptoms of various chronic disease that can be utilized for early diagnosis [13]. The alterations in vital signs are sometime so rapid that even doctors fail to correlate the causes quickly which can result in a serious clinical emergency, even death. Therefore, the development of techniques for early discovery of knowledge by utilizing intrinsic patterns in large quantities of vital sign data and scalable power of cloud computing [7] can enable doctors and care-givers to make accurate and real-time data-driven decisions. This will have an effective impact on patient-care, diagnosis and more importantly will reduce patient morbidity and mortality, and prevent outbreaks of illness and disease in hospitals.

Countries all over the world are facing the problem of chronic illness. According to US Centers of Disease Control (CDC) [14] as of 2012, 117 million people (about half of the total adult populations) have one or more chronic health conditions. Seven of the top 10 causes of death in 2010 were chronic illness [15]. Only heart disease and cancer together accounted for nearly 48% of all deaths [14]. Moreover, there are some other problems for irregular lifestyle and diet. About 78 million people were obese in 2010. Many people are also suffering diabetes which is a leading cause of kidney failure and blindness. The expenditure for treating these diseases in hospital is also very high. In 2010, the total costs of heart disease and stroke were estimated to be $315.4 billion and of this amount $193.4 billion was for direct medical expenses, not including hospital costs[14]. If chronically ill patients wait a long time for their next hospital visit, which is eventually a costly procedure, then a constant threat of uncertainty can make the chronic conditions even intense. Therefore, if all vital signs can be monitored in real-time with early prediction capability, then patients would have peace of mind and a higher chance of avoiding the next chronic episode. Thus, a proactive prevention system is a challenge for modern healthcare technology.

Generally in a cloud-based remote monitoring system, various health parameters of a patient are continuously collected from a number of sensors and delivered to the cloud through a network. Intelligent algorithms run in cloud servers to find any sign of abnormalities in those collected data and the outcomes are sent to the healthcare professionals for assessing a patient's health condition. The amount of clinical data that is accumulated per patient is enormous and growing. Such data are not only massive in size but also heterogeneous in nature due to the clinical essence and medical history of patients [10]. If a system remotely monitors all patients with a common chronic condition such as hypertension and the continuous data of all vital signs are stored in a cloud database, that data can be analysed to forecast when patient within that group is likely to have an abnormal clinical event such as a heart attack. If we just learn from the vital sign data of all such patients 2 h preceding a negative clinical event then we can utilize that knowledge in real-time to predict that clinical event 2 h earlier for a new unknown patient. As the data grow and learning algorithms improve, more patients will benefit from the collective knowledge gained from the data. The high speed network communications, high resourceful cloud platforms and readily available cloud services have simplified the process of transferring, gathering and analysing these big biomedical data.

1.1. Motivations

The demand for a preventive, protective and reliable healthcare services have increased the interest of developing home-based remote monitoring systems with intelligent decision support. One of the major issues in such system is the accurate and early detection of future abnormalities to protect patients from potential health-related threats. Including the challenge of building a proactive prevention system, we are also motivated by the findings of our earlier studies to proceed with this work. Previously, we had developed a personalised knowledge discovery framework named BDCaM [10] for assisted healthcare using context-aware big data and a cloud-based model, CoCaMAAL [7]. In the BDCaM model, we developed learning methods to discover patient-specific correlations using 2 vital signs (HR and BP), patient activity and environmental temperature which are then utilized to identify patient-specific future abnormalities. We ignored the correlations among vital signs which contain useful patterns for the progression of some chronic diseases. We first utilized multiple vital sign correlations in developing a real-time probabilistic predictor system using the Hidden Markov Model (HMM) for early discovery of clinical events [16]. There we used about 1023 patient records and a small number of features were extracted from the observed data just before the clinical event. In this work, more features are extracted from 1 to 2 h past observed data of the clinical event. Here, we also focus on clinical knowledge discovery based on interactions among multiple vital signs but using a large number of features obtained from recent past data.

We are also encouraged by some of our previous work in biomedical data analysis [17,18], disease diagnosis [19], abnormality prediction and change detection [20], wavelet transform [21,22], decision support [5], anomaly detection [23,24], wireless sensor network [25] and cloud-based remote monitoring [26] to develop this leaning model for early detection of abnormal clinical events in real-time personalized monitoring. An example of strong correlation between HR and BP is shown in Figs. 2 and 3.

There has been limited research attempting to forecast various clinical events using multi-parameter data of a large number of patients [27]. Most of those studies used a small sample of data (i.e. a few megabytes) from a small group of patients, short-length forecast window (mostly an hour) and work only on a single parameter such as blood pressure [28], ECG [17] or PPG. Those models utilized a small number of features for training and can only predict the future behaviours of some special clinical events [29,30]. Such systems suffer a high number of false alerts when uncertainty of data goes higher with the increase in patient numbers.

1.2. Contributions

Our goal is to build a predictive model that learns about various known medical conditions by mining historical multi-parameter data of many patients and discover patterns therein that can be used to make predictions for patients with unknown medical conditions. The early intervention capability would allow doctors to plan a preventive treatment, thus saving patients from possible

Download English Version:

https://daneshyari.com/en/article/4954751

Download Persian Version:

https://daneshyari.com/article/4954751

Daneshyari.com