Computers in Industry 97 (2018) 54–66

Contents lists available at ScienceDirect

Computers in Industry

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

# Data accuracy aware mobile healthcare applications

Lamia Ben Amor[a,*], Imene Lahyani[a,b], Mohamed Jmaiel[a,c]

[a] ReDCAD Laboratory, University of Sfax, National School of Engineers of Sfax, B.P. 1173, 3038 Sfax, Tunisia
[b] LAAS, CNRS, 7 avenue du Colonel Roche, F-31400 Toulouse, France
[c] Digital Research Center of Sfax, B.P. 275, Sakiet Ezzit, 3021 Sfax, Tunisia

ARTICLE INFO

*Article history:*
Received 31 May 2017
Received in revised form 25 November 2017
Accepted 31 January 2018
Available online xxx

*Keywords:*
Accuracy
Mobile healthcare application
Anomaly detection
Anomaly isolation
Dimension reduction
Principal Component Analysis
Robustness

ABSTRACT

This paper proposes a new approach for online detection and isolation of inaccurate vital sign measurements in mobile healthcare applications. Our primary objective is to distinguish between inaccurate measurement and patient health degradation to reduce false medical alarms. The proposed approach couples dimensionality reduction with inaccurate data detection and isolation. On the one hand, dimension reduction is based on robust incremental Principal Component Analysis. On the other hand, multivariate anomaly detection relies on squared prediction errors and anomalous vital sign isolation is based on contribution plots. We apply our proposed approach on real medical dataset. Our simulation results prove the effectiveness of our approach in achieving good recall with a low false alarm rate compared with existing solutions. The benefit gained by our approach in terms of time and space complexities make it useful and efficient for real time mobile healthcare applications.


## 1. Introduction

Mobile health [1,2] can be broadly defined as the use of information and communication technologies that are accessible to people or healthcare providers through mobile devices such as mobile phones, patient monitoring devices other wireless devices to support medical and public healthcare practices. mHealth plays a significant role in delivering remote healthcare services while empowering people to better manage their own health state and improve the ability to track diseases. It aims at reducing socio-economic costs [3] and even surmount organizational barriers [2].

Mobile healthcare systems are rapidly growing and evolving. The architecture of such system contains three domains. In the device domain, a patient is equipped with a set of wearable and/or implantable sensors that constantly sense vital signs. Sensors are interconnected via a medical body area network. The central point of the device domain is the application hosting device. It aggregates vital sign measurements and transmits them via the network domain to the application domain. Mobile health applications cover several areas to include disease management (such as chronic disease management [4], women's health and pregnancy [5] and elderly people monitoring [6]), as well as, wellness management [7].

Particularly, mobile health applications highly depend on continuous sensor readings to provide high-quality healthcare services and improve the quality of patient supervision [8]. However, sensor readings introduce critical accuracy challenges [9] due to several internal and external factors. Internal factors are related to the performance variation of sensors involved in the application such as energy depletion, hardware faults, calibration, etc., and other external environmental-related ones.

To achieve its full potential, a mHealth application needs to deal with medical data accuracy challenges.

In the last few years, there has been a growing interest in proposing approaches dealing with anomaly detection and isolation for mhealth applications. The proposed approaches are divided into two main classes including supervised [10–12] and unsupervised approaches [13–17].

Despite the tremendous research effort around approaches for anomaly detection and isolation in mHealth applications, there are still several issues preventing their deployment in real time healthcare systems. Classification-based supervised approaches [10,12] rely on trustable training dataset. However, this assumption may not be conventional in real time mHealth applications, where anomalies are not known in advance. Furthermore, training data often contains abnormal measurements which lead to overfitted classification or regression models. A majority of existing unsupervised approaches [14,13,16] cannot be applicable in high-dimensional data and can generate additional complexity in terms of time and space. Indeed, in real time mHealth applications, wearable devices such as smart vests [18] and other

* Corresponding author.
*E-mail addresses:* lamia.benamor@redcad.org (L. Ben Amor), iabdenna@laas.fr (I. Lahyani), jmaiel.mohamed@enis.rnu.tn (M. Jmaiel).

https://doi.org/10.1016/j.compind.2018.01.020
0166-3615/

implantable sensors like Implantable Cardioverter Defibrillator [19] increase the amount of multi-dimensional medical data by being able to send vital signs at any time and any place.

In this paper, we address these issues by proposing an unsupervised anomaly detection and isolation approach for mHealth applications. The proposed approach combines two key characteristics. First, it couples a dimension reduction algorithm based on the spatial and temporal correlation among monitored vital signs with anomaly detection and isolation methods. Second, it uses lightweight and robust unsupervised techniques in both steps. Dimension reduction is based on robust incremental principal component analysis which is not sensitive to anomalies and precludes the necessity of training using a trustable dataset, a strong advantage from the operational point of view. Multivariate anomaly detection is based on squared prediction errors and adequate threshold. Once multivariate anomaly is noticed, an isolation step is carried out using contribution plots [20] to identify suspicious vital signs.

Simulations in real medical record are firstly conducted to choose the suitable principal component analysis for dimensionality reduction. Second, our simulation results prove that our proposed approach achieves good detection performances and demonstrate the importance of dimension reduction in reducing computational time and memory. Finally, they validate that our approach outperforms other related approaches [11,13].

The rest of the paper is organized into four sections: Section 2 outlines related work. The proposed approach is discussed in Section 3. Section 4 shows the experimental results of this work. Finally, Section 5 includes a summary and gives some directions for the future.

## 2. Literature review

Various mobile healthcare applications for periodic vital signs monitoring have been proposed, such as CodeBlue [21], HealthGear [22], SmartVest [18], MEDiSN [23], PIaIMoS [24], and so on. Their main features are summarized in Table 1.

Recent surveys of mobile healthcare applications are available in [24,25].

Unfortunately, real-time vital sign measurements often contain inconsistencies and errors due to diverse internal and external factors. The internal factors are, especially, related to the performance of mobile devices involved in the application such as energy depletion [9], hardware failure, sensor calibration [8], and so on. External factors include some environmental-related ones. For instance, fluorescent lighting may cause data errors in the pulse oximeter. As sensor can be wearable, an incorrect placement of the device on patient's body, however, may sometimes lead to inaccurate measurements. Existing mobile Healthcare applications do not deal with data accuracy issues.

Quite recently, considerable attention has been paid to data anomaly detection and isolation in mobile healthcare applications. Existing approaches can be classified into two main categories, namely supervised and unsupervised approaches.

Supervised approaches involve classification [12,10] and regression [11] methods. In [12], authors have proposed an anomaly detection algorithm based on decision tree algorithm (J48) to classify new patient records as normal or abnormal. In fact, collected physiological parameters are represented by tree nodes and classes (normal and abnormal) are represented by the leaf nodes. Once an anomaly is detected, a regression algorithm is performed to distinguish between a faulty sensor reading and a patient anomaly. In [10], authors have described an anomaly detection and isolation approach based on Support Vector Machine (SVM) to classify abnormal observations in the incoming sensor readings. The objective is to construct a maximum margin separating hyperplane which divides the training data into two classes normal and abnormal. Each new physiological observation outside the hyperplane is classified as abnormal. If detected, the authors apply a regression algorithm to determine if the patient health degrades or if a sensor measurement is faulty.

In [11], authors have provided an univariate anomaly detection approach for healthcare applications based on three algorithms. The Sequential Minimal Optimization (SMO) Regression is used to predict a sensor value from historic values. The difference between the predicted and the actual sensor values is compared with a dynamic threshold. In fact, the threshold represents the standard deviation of the collected physiological measurements. The Majority Voting (VM) algorithm is used to decide whether the upcoming sensor value is anomalous. However, anomaly detection algorithms for mHealth applications must be able to analyze multivariate data and exploit the correlation among the physiological parameters. Thus, analysis of multivariate data improves the accuracy of anomaly detection techniques. Furthermore, historic data often contain anomalies which may affect the efficiency of the SMO regression algorithm. Thus, the need of robust regression techniques is of paramount importance.

**Table 1**
Features of mobile healthcare applications.

| Application | Medical sensor used | Communication system | Application hosting device |
|---|---|---|---|
| *CodeBlue* [21] | Blood oxygen sensor<br>ECG[1]<br>EMG[2] | Wireless sensor network | Computer |
| *HealthGear* [22] | Pulse oximeter | Bluetooth | Bluetooth-enabled cell phone |
| *SmartVest* [18] | Blood pressure<br>ECG<br>EMG<br>EEG[3]<br>Thermometer<br>Pulse oximeter | Wireless sensor network | Remote monitoring station |
| *MEDiSN* [23] | ECG<br>Pulse oximeter | Wireless sensor network | Not provided |
| *PIaIMoS* [24] | ECG<br>Blood oxygen sensor<br>Respiration rate sensor<br>Galvanic sensor | Wireless sensor network | Computer tablets<br>Smartphone |

[1] Electrocardiogram.
[2] Electromyography.
[3] Electroencephalogram.