Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 108C (2017) 99–108

Procedia Computer Science

International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland

Anomaly Detection in Clinical Data of Patients Undergoing Heart Surgery

Alva Presbitero[1*], Rick Quax[2], Valeria Krzhizhanovskaya[1,2] and Peter Sloot[1,2,3]
[1]*ITMO University, St. Petersburg.*
[2]*University of Amsterdam, The Netherlands*
[3]*Nanyang Technological University, Singapore*
[*]*avpresbitero@gmail.com, r.quax@uva.nl, V.Krzhizhanovskaya@uva.nl, p.m.a.sloot@uva.nl*

Abstract

We describe two approaches to detecting anomalies in time series of multi-parameter clinical data: (1) *metric and model-based indicators* and (2) *information surprise*. (1) *Metric and model-based indicators* are commonly used as early warning signals to detect transitions between alternate states based on individual time series. Here we explore the applicability of existing indicators to distinguish critical (anomalies) from non-critical conditions in patients undergoing cardiac surgery, based on a small anonymized clinical trial dataset. We find that a combination of time-varying autoregressive model, kurtosis, and skewness indicators correctly distinguished critical from non-critical patients in 5 out of 36 blood parameters at a window size of 0.3 (average of 37 hours) or higher. (2) *Information surprise* quantifies how the progression of one patient's condition differs from that of rest of the population based on the cross-section of time series. With the maximum surprise and slope features we detect all critical patients at the 0.05 significance level. Moreover we show that a naive outlier detection does not work, demonstrating the need for the more sophisticated approaches explored here. Our preliminary results suggest that future developments in early warning systems for patient condition monitoring may predict the onset of critical transition and allow medical intervention preventing patient death. Further method development is needed to avoid overfitting and spurious results, and verification on large clinical datasets.

Keywords: Anomaly Detection, Information Surprise, Early Warning Signal, Time series Analysis

1 Introduction

Anomaly detection is the process of pinpointing and segregating items in a population exhibiting behaviors that deviate from the norm. These are referred to as “anomalies” or “outliers”. Anomaly detection is extensively used in detecting fraud credit cards [1], cyber security [2, 3], health insurance [4], and patient monitoring using electrocardiography (ECG) signals [5]. For instance, technologies in

1877-0509 © 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the scientific committee of the International Conference on Computational Science
10.1016/j.procs.2017.05.002

anomaly detection, especially for medical applications, are essential for estimating physical conditions or states of patients from health to demise. Often times, when anomalies occur, significant changes in time series patterns become evident. These anomalies (critical conditions) designated by pattern changes could serve as indicators of transitions from healthy to critical state that leads to death in patients.

Numerous complex dynamical systems have been found to exhibit transitions or tipping points where systems abruptly shift from one stable state to another. The specific case of disease can be regarded as a sudden shift in system state from health to disease [6, 7]. For instance, the onset of depression is explored by looking at fluctuations of emotions as indicators of transition from a normal to a state of depression [8]. Other examples include financial systems, which exhibit systemic market crashes [9, 10], climate shifts preceded by the slowing down of fluctuations [11–13], decline of population leading to extinction [14–16], flood early warning systems [17–19] and dams [20].

Early warning signals (EWS) are used as indicators for loss of system resilience prior to transitions based on more subtle statistical properties of the measurements [21]. This can sometimes be detected through changes in correlations, standard deviation, or skewness of the series through time [22]. We utilize indicators used in EWS to segregate critical from non-critical patients with the assumption that critical patients exhibit pattern changes in their time series when anomalies or transitions from health to demise occur. We do not detect how much time these transitions occur in advance. We aim to incorporate this in the future version of our work.

In the present work we explore the applicability of using four classical EWSs on blood parameter concentration time series from 53 patients undergoing complicated cardiac surgery to detect the transitions of patient death; an approach that has not been done in literature before. The most important motivation of using EWS is its potential of real-time use as early warning for increased risk of patient death, with eventually the goal of improved prevention.

2 Data Preparation and Analysis

The raw data consists of concentrations of 36 different blood parameters from 53 patients (total of 878 sample data points) undergoing complicated cardiac surgery (including timestamps), three of whom died after the operation [23]. Patients are composed of male or non-pregnant, non-lactating female of any race with an age over 18. The period to which the data was collected comprised a time interval of 24 hours prior to the time of surgery up until 30 days after surgery. Several blood samples were usually taken within 24 hours after the surgery as this time period is the most critical for patients who have recently undergone surgery. Patients then usually stabilize after this point so the rate at which blood is sampled is reduced.

Raw data also contains a substantial fraction of missing data points of 62.5% because not all 36 parameters were always tested for in all blood samples. However the missing values are reasonably well distributed over the parameters: the 95% confidence interval (CI) of missing values is 43.8%-88.8%. Missing data will be inevitable in clinical trial data so any EWS method must be capable of dealing with it. There are various techniques to deal with missing values but most importantly the technique should not significantly increase the rate of false negatives (labeling healthy patients as critical) because this would render the signal noisy and make it impractical for medical practitioners to act upon it.

Figure 1 shows a sample of the bootstrapped raw data for IL10, one of the blood parameters in the raw data. Red corresponds to the critical patients while blue corresponds to non-critical patients. Increasing concentrations of IL10, a type of anti-inflammatory cytokine that aids the human body against foreign attacks, in patients usually indicate the presence of inflammation. We see an increasing concentration for critical patients. However, we note that we only have three samples for the critical patients, hence bootstrapping is only limited to at most three points per time step.