Journal of Biomedical Informatics 53 (2015) 81–92

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Integrating monitor alarms with laboratory test results to enhance patient deterioration prediction

Yong Bai [a], Duc H. Do [b], Patricia Rae Eileen Harris [c], Daniel Schindler [c], Noel G. Boyle [b], Barbara J. Drew [c], Xiao Hu [c,d,e,f,*]

[a] Department of Bioengineering, University of California, Los Angeles, CA, United States
[b] UCLA Cardiac Arrhythmia Center, David Geffen School of Medicine, University of California, Los Angeles, CA, United States
[c] Department of Physiological Nursing, University of California, San Francisco, CA, United States
[d] Department of Neurosurgery, University of California, San Francisco, CA, United States
[e] Institute for Computational Health Sciences, University of California, San Francisco, CA, United States
[f] UCB/UCSF Graduate Group in Bioengineering, University of California, San Francisco, CA, United States

ARTICLE INFO

Article history:
Received 10 March 2014
Accepted 9 September 2014
Available online 18 September 2014

Keywords:
Alarm fatigue
Monitor alarm
Maximal frequent itemsets mining
Clinical deterioration
Code blue
Event prediction

ABSTRACT

Patient monitors in modern hospitals have become ubiquitous but they generate an excessive number of false alarms causing alarm fatigue. Our previous work showed that combinations of frequently co-occurring monitor alarms, called SuperAlarm patterns, were capable of predicting in-hospital code blue events at a lower alarm frequency. In the present study, we extend the conceptual domain of a SuperAlarm to incorporate laboratory test results along with monitor alarms so as to build an integrated data set to mine SuperAlarm patterns. We propose two approaches to integrate monitor alarms with laboratory test results and use a maximal frequent itemsets mining algorithm to find SuperAlarm patterns. Under an acceptable false positive rate FPR_{max}, optimal parameters including the minimum support threshold and the length of time window for the algorithm to find the combinations of monitor alarms and laboratory test results are determined based on a 10-fold cross-validation set. SuperAlarm candidates are generated under these optimal parameters. The final SuperAlarm patterns are obtained by further removing the candidates with false positive rate > FPR_{max}. The performance of SuperAlarm patterns are assessed using an independent test data set. First, we calculate the sensitivity with respect to prediction window and the sensitivity with respect to lead time. Second, we calculate the false SuperAlarm ratio (ratio of the hourly number of SuperAlarm triggers for control patients to that of the monitor alarms, or that of regular monitor alarms plus laboratory test results if the SuperAlarm patterns contain laboratory test results) and the work-up to detection ratio, WDR (ratio of the number of patients triggering any SuperAlarm patterns to that of code blue patients triggering any SuperAlarm patterns). The experiment results demonstrate that when varying FPR_{max} between 0.02 and 0.15, the SuperAlarm patterns composed of monitor alarms along with the last two laboratory test results are triggered at least once for [56.7–93.3%] of code blue patients within an 1-h prediction window before code blue events and for [43.3–90.0%] of code blue patients at least 1-h ahead of code blue events. However, the hourly number of these SuperAlarm patterns occurring in control patients is only [2.0–14.8%] of that of regular monitor alarms with WDR varying between 2.1 and 6.5 in a 12-h window. For a given FPR_{max} threshold, the SuperAlarm set generated from the integrated data set has higher sensitivity and lower WDR than the SuperAlarm set generated from the regular monitor alarm data set. In addition, the McNemar's test also shows that the performance of the SuperAlarm set from the integrated data set is significantly different from that of the SuperAlarm set from the regular monitor alarm data set. We therefore conclude that the SuperAlarm patterns generated from the integrated data set are better at predicting code blue events.

* Corresponding author at: Departments of Physiological Nursing/Neurosurgery, Institute for Computational Health Sciences, University of California, San Francisco, 2 Koret Way, N611J, San Francisco, CA 94143-0610, United States.
E-mail address: Xiao.Hu@nursing.ucsf.edu (X. Hu).

http://dx.doi.org/10.1016/j.jbi.2014.09.006
1532-0464/

1. Introduction

With technologic advances in medical devices over the past few decades, life-saving patient monitoring systems have become ubiquitous in modern hospitals [1]. Alarms annunciated by the monitoring systems are expected to alert caregivers to either changes in monitored physiological parameters of a patient or device malfunction, and to enhance quality of care and patient safety by detection of any abnormality [2].

In traditional monitor algorithms, an alarm is triggered immediately when the value of the monitored parameter exceeds or falls below the preset threshold [3]. Due to the lack of a standard for default threshold setting [4], this threshold-based algorithm is intentionally set to have high sensitivity in order to capture the greatest percentage of clinically significant events [5,6]. As a consequence, there is low specificity and numerous alarms occur (about 700 alarms per patient per day [7]) and up to 99% of them are false alarms and nuisance (or false positive) alarms with no clinical relevance [2,5,7–10]. Caregivers exposed to a large number of false and nuisance alarms become desensitized, leading to alarm fatigue problems [7,9,11]. Excessive false and nuisance alarms may compromise the quality of patient care and cause unexpected alarm-related deaths in hospitals [12]. The alarm hazard has been ranked as the "TOP 1" technology hazard for 2014 by the Emergency Care Research Institute (ECRI) [13].

Many studies have focused on addressing the alarm fatigue problem. Descriptions of many such algorithms were provided in reviews [1,14]. For instance, Zong et al. [15] proposed an algorithm for reducing false arterial blood pressure (ABP) alarms by evaluating signal quality of ABP and the relationship between electrocardiogram (ECG) and ABP using fuzzy logic approach. Similarly, Aboukhalil et al. [16] reduced false critical ECG arrhythmia alarms using morphological and timing information derived from the ABP waveforms. Lastly, Li et al. [17] used a machine learning technique and data fusion method to reduce false arrhythmia alarms by combining signal quality and physiological metrics derived from the waveforms of ECG, photoplethysmograph, and optionally, ABP. We applied pattern recognition methods to reduce false intracranial pressure (ICP) alarms using the morphological waveform features extracted from the ICP signal [18,19]. These approaches were developed to manage individual alarm types and further validation is needed to ensure that no true alarm is suppressed before their implementations by monitor vendors. Additionally, true alarms not suppressed by these approaches were designed to detect abnormalities after they occur, not to detect patient deterioration. Therefore, they are at best able to support a reactive patient care practice rather than a predictive one.

To detect patient deterioration, especially outside intensive care units, several score-based systems have been developed based on multiple parameters. The modified early warning score (MEWS) [20], for instance, was a simple tool to produce a fusion score based on the summation of an individual score assigned to each of five physiological parameters: systolic blood pressure (SysBP), respiratory rate (RR), pulse rate, temperature and patient consciousness. For each parameter, the greater the degree of deviation from the normal range, the larger the individual score assigned. However, the schema for score assignment was designed empirically [21]. Biosign [22,23] was another algorithm to generate a patient status index (PSI) by fusing five vital signs: heart rate (HR), respiratory rate (RR), blood pressure (BP), temperature and arterial oxygen saturation (SpO2). It used a multivariate Gaussian probabilistic model for the distribution of these vital signs for patients without crisis events. A patient crisis event was detected when these vital signs had a small probability according to this distribution estimated from a training data set. Rothman et al. [24] developed a system to calculate a patient acuity metric, called the Rothman Index (RI), to evaluate the risk of patient deterioration using vital signs, laboratory test results, indicators of cardiac rhythms, and nursing assessments. This approach was based on empirical accumulation of relative risks of its component variables in determining patient mortality after 1 year discharge from the hospital. Machine learning-based methods have also been proposed to detect patient deterioration. For instance, Clifton et al. [25] compared Gaussian mixture model (GMM) and support vector machine (SVM) with HR, RR, SpO2, and SysBP as input. Tarassenko et al. [26] developed a centile-based early warning score system based on statistical properties of the vital signs (HR, RR, SpO2 and SysBP) to identify deteriorating patients. Scores were determined when the statistical value of vital sign fell into certain range of centile.

It can be argued that those algorithms presented above for detection of patient deterioration introduce additional alarms or alerts without providing direct relief of the existing alarm fatigue problem. A potentially more desirable approach would incorporate patient monitor alarms and physiological signals from patient monitors. The idea to include monitor alarms as predictors of patient deterioration detection models has been tested by our group. In our previous paper [27], we proposed a novel data-driven approach using raw streaming alarm data to: (1) identify patterns that were combined with different monitor alarms using in-hospital code blue events; (2) select those patterns that occurred sufficiently often preceding code blue events but rarely in control patients; (3) empirically define and determine the optimal length of time window for the selected patterns; (4) assess the temporal characteristics of these patterns such as the sensitivity with respect to prediction window; and then (5) based on these factors, evaluate the performance of these patterns, which we called SuperAlarm patterns, under varying acceptable false positive rates. Because a SuperAlarm trigger necessarily requires simultaneous triggering of different alarms, it therefore has the potential to reduce alarm frequency.

In the present study, we follow the general framework we have previously proposed [27] and describe how we extend the conceptual domain of a SuperAlarm to incorporate laboratory test results as an additional source to compose SuperAlarm patterns. To do so, we propose several new methods so as to tackle complicating factors that arise when one incorporate non-streaming data (e.g., patients with very sparse data). We also address the need to exclude "crisis" alarms that clinicians would consider to be "no brainers" such as asystole. Specifically, we first explore a Non-Homogenous Poisson Process (NHPP) to model the occurrence rate of monitor alarms and obtain an objective threshold to exclude code blue patients with unexpectedly small number of monitor alarms preceding code blue events. We then develop two approaches to integrate laboratory test results with monitor alarms. We apply a new algorithm to discover SuperAlarm candidate patterns occurring frequently before code blue events. These candidate patterns are composed of combinations of maximal number of monitor alarms and laboratory test results with occurrence rate greater than a support threshold. The candidate patterns are further filtered out if their false positive rates are greater than an acceptable false positive rate FPR_{max}, resulting in the final SuperAlarm patterns. By construction, these patterns are less redundant compared to those determined by the techniques of mining frequent itemsets (FI) or closed frequent itemsets (CFI) used in our previous work.

2. Methods

Fig. 1 illustrates the flowchart of the proposed algorithm to discover SuperAlarm patterns. Key steps of this process are described in the following sections.