Proceedings of the 29th Annual International
Conference of the IEEE EMBS
Cité Internationale, Lyon, France
August 23-26, 2007.

Real-Time Evaluation of Patient Monitoring Algorithms for Critical Care at the Bedside

Ying Zhang, MEng, Christine Tsien Silvers, MD, PhD, and Adrienne G. Randolph, MD, MSc

Abstract—**Rapid interpretation of physiological time-series data and accurate assessment of patient state are crucial to patient monitoring in critical care. Algorithms that use artificial intelligence techniques have the potential to help achieve these tasks, but their development requires well-annotated patient data. In this study, we designed a data acquisition system for synchronized collection of physiological time-series data and clinical event annotations at the bedside to support the evaluation of alarm algorithms in real time, and implemented this system in a pediatric intensive care unit (ICU). This system captured vital sign measurements at 1 Hz and 325 clinical alarms generated by the bedside monitor and the 2 instances of false negatives during a monitoring period of 196 hours. The alarm annotations in real time at the bedside indicate that about 89% of these alarms were clinically-relevant true positives; 6% were true positives without clinical relevance; and 5% were false positives. These findings show an improved specificity of the alarm algorithms in the newer generation of bedside monitoring systems and demonstrate that the designed data acquisition system enables real-time evaluation of patient monitoring algorithms for critical care.**

I. Introduction

While more physiological time-series data are available than ever before in critical care, how to integrate, represent, and utilize them in real time has remained a challenging clinical and engineering question. "Intelligent" computer algorithms that recognize patterns in the association between physiological time-series data and clinical events could shed light on this question. Crucial to the development of such algorithms are real patient data and a way to reconstruct the clinical context under which these data are generated.

The reconstruction of clinical context, however, is a nontrivial task. In previous studies, data acquisition in the ICU focused primarily on collecting physiological signals from the bedside monitors and clinical data from patients' medical records or, if at the bedside, using a separate acquisition system [1]. Experts then retrospectively analyzed these data to derive the clinical context. Little information about the state of the patient and clinical events were recorded at the bedside as they occurred. This approach has several limitations. First, because physiological data and clinical annotations are collected separately, the two datasets can be poorly synchronized. Second, physiological data and clinical annotations have different time "granularity;" this can make correlation of the data ambiguous since it is difficult retrospectively to determine the timing of a clinical event down to seconds. Third, in the ICU setting, it has not been feasible to record everything that might be useful retrospectively; thus, clinical annotations are collected based on projected future research needs. Retrospectively, however, additional clinical information is often needed. Reconstructing the clinical context with these limitations consequently results in significant uncertainties and introduces assumptions that can make subsequent results largely speculative [1].

To help obtain the true clinical context, this study introduces a system that enables synchronized collection of physiological data with bedside clinical event annotation. This system has been implemented in a pediatric ICU and has been used to evaluate both the bedside monitor's alarm algorithms and "intelligent" alarm algorithms in real time.

II. Methods

A. Physiological Data Collection

The Physiological Data Collection Unit is an automated module of the data acquisition system residing on a standard laptop controlled by a trained observer at the bedside. It communicates with the HP Viridian Neonatal Component Monitoring System (CMS), the bedside monitor, via an RS232 interface (Option 13 for CMS Model 1077A). Its core is a 16-bit application written in Turbo C based on the source code of CMS's interface demo program, CMSCOM [2].

Collecting all physiological data that the monitor could provide would be ideal; however, the maximum bandwidth of the communication channel is 38,400 baud. Due to this limitation, the Physiological Data Collection Unit is usually set to collect the numeric parameters (e.g., heart rate, blood pressure, arterial oxygen saturation) that are consistently and frequently monitored. Data granularity is at 1 Hz, the highest frequency at which new values become available.

Manuscript received June 22, 2007. This work was supported in part by DARPA Grant F30602-99-0509, a Harvard-MIT Health Sciences and Technology MEMP Fellowship, and an NLM Medical Informatics Training Grant.

Y. Zhang is with Harvard Medical School-MIT Division of Health Sciences and Technology and MIT Department of Electrical Engineering and Computer Science, Cambridge, MA 02139 USA (phone: 617-253-4679; fax: 617-258-8682; e-mail: yingz@mit.edu).

C. Silvers is an Attending Physician, Department of Emergency Medicine, at Caritas Good Samaritan Medical Center, Brockton, MA 02301 USA; Jordan Hospital, Plymouth, MA 02360 USA; and Massachusetts General Hospital, Boston, MA 02114 USA. She is also a Research Affiliate, Children's Hospital Informatics Program, Boston, MA 02115 USA (e-mail: tsien@alum.mit.edu).

A. Randolph is an Associate Professor of Anaesthesia, Harvard Medical School, Boston, MA 02115 USA (e-mail: Adrienne.Randolph@childrens.harvard.edu).

B. Command Center

The Command Center controls data transfer, data synchronization, and all user-selected modules through the main user interface. The Command Center is written in Visual Basic .NET for the ease of creating and modifying the user interface, availability of multiple timers, and direct access to the SQL server and other database engines using the functionality ADO.NET.

The Command Center obtains physiological data from data text files and transfers them into the database at the selected clock times. This clock is based on the system time of the laptop. The Command Center uses this clock to check for newly arrived physiological data in the past two seconds. It then time-stamps each set of new data with the time of this clock and records the data into the database. If no new data has arrived for a particular numeric parameter, the Command Center automatically records a number designated for missing values along with the new values of other numerics.

C. Clinical Event Annotation

Clinical event recording is done in four cases: 1) the bedside monitor sounds an alarm; 2) an algorithm under investigation displays an alarm; 3) the patient becomes agitated and requires immediate attention when no alarm occurs; and 4) medications are being administered or discontinued. When an event occurs, an annotation box, as in Figure 1, is activated by alarm information from the bedside monitor in case 1, by the algorithm in case 2, or by a trained observer in cases 3 and 4.

The fields "Begin Time," "Alarm Type," and "Alarm Severity" are automatically filled when the annotation box appears. "Begin Time" is the time point when the alarm starts. "Alarm Type" is the physiological parameter that triggers the alarm. "Alarm Severity" is given by the monitor. "End Time" is automatically filled in when the alarm stops.

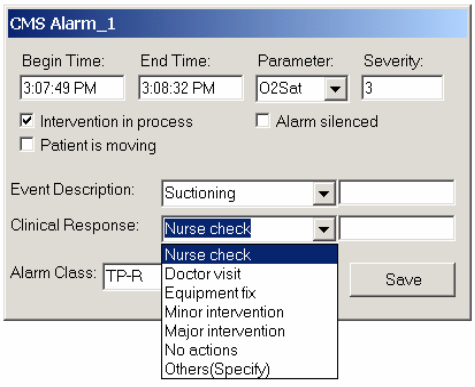

Fig. 1. An example of the annotation box.

Technical alerts generated by the bedside monitor are denoted as "INOP"s. These are triggered by signal quality problems, equipment malfunction, measurement setup problems, or ongoing calibration [3]. INOPs are recorded into the database but are not annotated by the trained observer because their causes are non-clinical.

During or after an alarm, the bedside trained observer records whether the patient is moving or whether a medical procedure is in progress, as well as how the medical staff responds to the alarm. He or she asks the nurse or physician at the bedside to classify the alarm into one of three categories, as adopted from work by Tsien [4]: TP-R (True Positive, Clinically Relevant), TP-I (True Positive, Clinically Irrelevant), FP (False Positive).

D. Database

The database is designed using SQL server desktop version for its ease in backing up and restoring different databases and, more importantly, for its enforced transactions mechanism, which ensures the physical integrity of each transaction. These features safeguard data integrity and prevent partial or corrupted records from being stored.

Time synchronization between different physiological parameters is achieved by aligning their time stamps. Time synchronization between event recordings is achieved by comparing their begin times and end times. Synchronization between physiological data and event recordings is achieved by going over all the time intervals defined by the begin time and end time of each event and identifying those events whose time intervals encompass the time stamp of each physiological data point.

E. Multiple Threading

Clinically-relevant events tend to cluster in time. They might be concurrent, overlapping, or in series. During such times, the caregivers are especially occupied such that they might not be available or have adequate information to classify each alarm. Our system allows multiple alarm messages to coexist and multiple annotation boxes to remain open in order to accommodate the annotation process. This is accomplished via multiple threading.

The primary thread serves the main user interface, initializes new threads, and relays commands to other threads in the system. The main thread has a timer clocking at 1 Hz to automatically read, process, and store the physiological data. When an alarm is generated, a new thread is initialized and triggers an annotation box to appear on the screen. This new thread will remain active until the user finishes entering the annotations about the alarm and closes the annotation box. Thus, concurrent or overlapping events can be annotated as the necessary clinical information becomes available.

F. Gold Standard for Algorithm Evaluation

The most difficult part of annotating an alarm-sounding event in the ICU is the accurate classification of the alarm. Even if the definitions of the alarm classes TP-R, TP-I, or FP are clearly disjoint, the complexity of the event or how an alarm is classified, by whom, and on what basis could make the classification process ambiguous. Currently, no reliable gold standard for alarm classification exists in the literature. In this study, we construct a 2-tier gold standard:

we first ask the patient's nurse or physician to classify an alarm-sounding event. This classification can then be revised within 30 minutes since subsequent events during this "grace period" could change the nurse's or physician's initial view of the patient state. We thus use the patient's condition in the *future* (the next 30 minutes) to revise the human expert's classification as needed.

G. System Evaluation

Evaluation of the designed data acquisition system was done in three stages: simulation, implementation, and usage. During simulation, both unit-level testing and system-level testing were performed. System-level testing was performed after each unit was shown to function properly. The tested system was then used to run alarm algorithms in real time in a completely functional phase, facilitating "stress" testing.

This system has been implemented and used at the Multidisciplinary Intensive Care Unit (MICU) of Children's Hospital in Boston. The study was approved as a part of a research protocol by the Institutional Review Board of Children's Hospital. A patient consent form was required to ensure that patients and their families were comfortable with the presence of the trained observer. Patient confidentiality and privacy have been protected according to hospital guidelines.

III. Results

The data acquisition system for synchronized collection of physiological time-series data and clinical event annotations at the bedside was tested for over 300 hours in sessions lasting from 2 to 12 hours in duration. Testing sessions in the later 196 hours demonstrated reliable operation. CMSCOM was the most robust and independent component of the system. Its function was not disrupted by disturbances or faults in other parts of the system. The number of concurrent multiple threads was tested up to 70.

The system's performance for data acquisition was consistent. For the evaluation of alarm algorithms, the system performed at a normal level for up to 10 algorithms. When trials of more than 10 alarm algorithms were carried out simultaneously, the high amount of computation and memory usage resulted in lower performance and data loss. Such problems were alleviated by decreasing the rate of data collection and analysis (e.g., from 1 Hz to 0.5 Hz).

Different patients were monitored for different sets of physiological parameters. Some parameters (e.g., heart rate) were measured in every patient, while other parameters were measured in only some patients. Table 1 lists the numeric parameters according to how frequently they were measured during the study period.

Sixteen patients participated in this study when the designed system was used to evaluate alarm algorithms. Five of these patients were so critically ill that the study sessions had to end within the first two hours. These patients had the highest alarm rates, above 10 alarms per hour, yet no INOPs were generated during their observation. The other eleven patients were followed for multiple sessions of 2-12 hours each, with an overall total of 196 monitoring hours.

The rates of the two types of alerts varied widely from session to session and patient to patient. Figure 2 is a histogram showing the number of patients in each bracket of different alarm rates.

Always monitored parameters	Heart rate (from electrocardiogram) Pulse rate (from pulse oximeter) Respiratory rate Arterial oxygen saturation
Frequently monitored parameters	Arterial line blood pressure (systolic, diastolic, mean) Noninvasive blood pressure (systolic, diastolic, mean) Oxygen perfusion Venous oxygen saturation
Less frequently monitored parameters	Temperature Central venous pressure Carbon dioxide level
Rarely monitored parameters	Wedge pressure Cardiac output Temperature difference

Table 1. Monitored numeric parameters

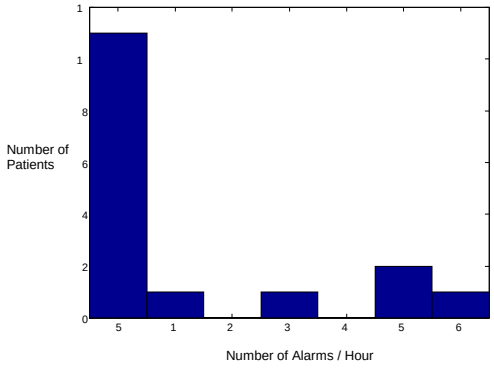

Fig. 2. Distribution of alarm rates for all patients

During the 196 monitoring hours, the bedside monitor sounded 325 clinical alarms; of these, 290 were true positives with clinical relevance, 20 were true positives but clinically irrelevant, and 15 were false positives. Two instances of false negatives were observed. The bedside monitor also generated 1768 INOPs. Unlike alarms, INOPs were not specifically indicative of a patient's condition.

Figure 3 depicts a histogram of the number of patients in each bracket of different alarm rates for the eleven patients. It is in fact a zoomed-in view of Figure 2 for the patients whose alarm rates were below 10 alarms per hour. Three of these eleven patients had no alarms. Others' alarm rates were relatively evenly distributed close to the mean of 1.7 alarms/hour. The total 1768 observed INOPs were randomly distributed among these patients.

IV. Discussion

This system for synchronized collection of physiological data and clinical annotations has been shown to correlate the physiological data and clinical event recordings in a consistent manner. Although its performance could be

influenced by hardware capabilities, the system achieved real-time data collection and evaluation of patient monitoring algorithms at the bedside.

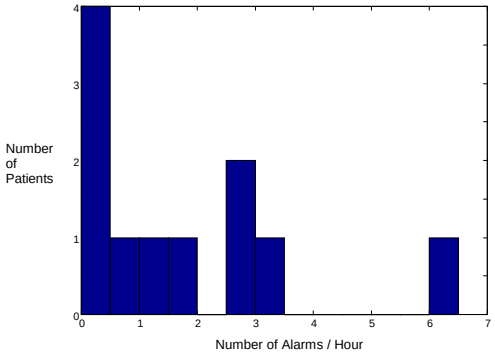

Fig. 3. Distribution of alarm rates for patients monitored for 2-12 hour sessions.

A major new finding in this study is the unexpectedly low volume of clinical alarms generated by the bedside monitor. In the study by Tsien [5], there were 2942 alarms during 298 monitored hours, i.e., about 9.9 alarms/hour. In a similar study by Lawless [6], there were 2176 alarms during 928 monitored hours, i.e., about 2.3 alarms/hour. Our results show only 325 clinical alarms during 196 monitored hours, or about 1.7 alarms/hour. These differences may come from 1) the identification of hardware or operational malfunctions as INOPs instead of clinical alarms, as illustrated in Table 2; 2) differences among patient populations; 3) improvements in biosensors and signal processing; and 4) more disease-specific algorithms of the newer generation of monitors.

Study	Average Alarm Rate (Number of Alarms / Hour)			
	Total	TP-R	TP-I	FP
Tsien	9.87	0.79	0.59	8.49
Lawless	2.34	0.13	0.62	1.59
This study (without INOPs)	1.66	1.48	0.10	0.08
This study (with INOPs)	10.68	1.47	0.11	9.10

Table 2. The frequencies of different types of alerts

While Tsien's study and this study were carried out at the same pediatric intensive care unit, the data collection systems and the bedside monitors were different. The patients monitored within the two studies, moreover, were different patients, who could be extremely different from one another in terms of their clinical conditions. Thus, it is important to consider variation in patient populations, and variation within a given patient population, especially given the limited number of patients available to this study and the resulting reduction in statistical validity and power.

Table 2 also shows that the rate of clinically-irrelevant true-positive alarms is lower than that in the previous studies by about sixfold. This decrease is further illustrated by the comparison of the TP-I alarms in Table 3. The significant reduction in the clinically-irrelevant true alarms is not the only evidence for improvement of new monitors: without counting INOPs, about 89.2% of all alarms are true positives with clinical relevance. Even with the INOPs counted (and thus the false alarm rate being comparable to those in the previous studies), the percentage of clinically-significant true alarms under the same classification method has improved from 8% to 13.8%.

Study	Percentage of Total Number of Alarms in Each Class		
	TP-R	TP-I	FP
Tsien	8%	6%	86%
Lawless	5.5%	26.5%	68%
This study (without INOPs)	89.2%	6.2%	4.6%
This study (with INOPs)	13.8%	1.0%	85.2%

Table 3. Class distribution of the alarms

We hypothesize that more sophisticated biosensors and signal processing methods have helped reduce noise and derive more accurate numeric values. Specialized algorithms, such as HP's STAR algorithm for detecting ST-segment elevation, may also have contributed to the increased specificity of alarm-sounding decisions [7].

In conclusion, the findings in this study suggest an improved specificity of new alarm algorithms used by newer generation bedside monitoring systems. The process of obtaining these findings has demonstrated that the designed data acquisition system enables real-time bedside evaluation of patient monitoring alarm algorithms for critical care. This system has facilitated accurate collection and correlation of physiological time-series data and clinical annotations, as well as prospective use of the clinical context in developing better patient monitoring algorithms.

ACKNOWLEDGMENT

The authors would like to thank Isaac Kohane, David Martin, Peter Szolovits, the entire MICU staff at Boston Children's Hospital, and all the patients and their families who kindly participated in this study.

REFERENCES

[1] Y. Zhang, "Real-time analysis of physiological data and development of alarm algorithms for patient monitoring in the intensive care unit," M.Eng. dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 2003.

[2] *HP Viridia Component Monitoring System RS232 Computer Interface Programming Guide*, 4th ed, Hewlett Packard Co., Palo Alto, CA, 1998.

[3] *Agilent M1177A Component Monitoring System User's Reference Manual*, 1st ed, Agilent Technologies, Andover, MA, 2000.

[4] C.L. Tsien, "TrendFinder: Automated detection of alarmable trends," *Laboratory for Computer Science Technical Report 809*, Massachusetts Institute of Technology, Cambridge, MA, July 2000.

[5] C.L. Tsien and J.C. Fackler, "Poor prognosis for existing monitors in the intensive care unit," *Crit. Care Med.*, vol. 25, Apr. 1997, pp. 614-619.

[6] S.T. Lawless, "Crying wolf: False alarms in a pediatric intensive care unit." *Crit. Care Med.*, vol. 22, June 1994, pp. 981-985.

[7] J. Wang, Agilent Technologies Healthcare Solutions Group, Andover, MA, private communication, July 2000.