Project Report - Alarm Fatigue

Weston Jackson

Advisors: Ansaf Salleb-Aouissi, Albert Boulanger

May 2017

1 Introduction

The CCLS has an ongoing project to explore and investigate the issue of alarm fatigue in physiological monitoring systems in hospitals, and to approach the problem of alarm fatigue from the perspective of machine learning. Alarm fatigue occurs when people are exposed to an excessive number of alarms, often resulting in desensitization to alarms and missed alarms [1]. While alarm fatigue has been studied in a very wide context, including nuclear control settings, the problem is especially relevant in the medical domain, where estimates state that over 72%-98% of all alarms are actually false.

There have been occasions when alarm fatigue was attributed as a direct cause of patient's death [2]. A survey by Healthcare Technology Foundation found that one in five respondent hospitals identified an avoidable adverse event due to alarms in the past two years [3]. ECRI Institute has ranked alarm fatigue as one of the top patient safety issue in its "Top 10 Technology Hazards" list ever since the list's inception [4].

2 Data Sets

2.1 Alarm Data

Columbia Neuro-ICU data set was extensively used in this project. The data set consisted of 489 patient admissions spanning from 2009-2013 in the Columbia Neuro-ICU. It consists of roughly 1 million alarms, 5,000 patient days, and 7 types of complications.

In a 2014 study, [5] identified three main types of alarms in intensive care units: equipment alarms, parameter alarms, and crisis alarms. We add an additional category, advisory alarms, in order to distinguish between the most life-threatening alarms and those that are less dangerous.

Crisis alarms include ASYSTOLE, VFIB/VTAC, VTACH, and VBRADY alarms. In total there were 1929 crisis events, consisting of approximately 0.2% of the data. Advisory alarms include other actionable alarms that do not have Hi/Lo parameter settings, such as BRADY, APNEA, VT $>$ 2, R ON T, TRIGEMINY, COUPLET, BIGEMINY, PAUSE, TACHY, PVC, and ARRHY SUSPEND. These were 110135 advisory events, consisting of roughly 11.5% of the data. Parameter alarms are Hi/Lo alarms that go off when oxygen, heart rate, blood pressure, and other leads are too high or low. There were 700313 parameter events, consisting of 73.3% of the data set. Finally, equipment alarms signify malfunctioning technical equipment in the intensive care unit. There were 127665 equipment events, consisting of roughly 13.4% of the data set. Approximately 1.6% of the alarms went unclassified.

2.2 Physiological Data

In addition to the alarm data, we also used physiological data from the NICU in order to corroborate this data set. We had physiological data for 102 of the 489 patients, which consisted mostly of heart rate levels. These levels are measured in five second intervals during the patient's entire stay at the ICU. Unfortunately, much of the data set is empty, and improving these physiological records should be a priority in the future.

Joining the physiological data and alarm data allows us to associate alarms with specific heart rate levels. In particular, there were 337479 events that we were able to match with physiological data. We provide confidence intervals for the mean heart rate of heart related crisis, advisable, and parameter alarms.

Alarm	Count	HR 95% CI
ASYSTOLE	303	(38.61, 45.67)
VFIB/VTAC	43	(72.13, 111.77)
BRADY	8354	(50.27, 50.51)
APNEA	877	(80.98, 82.91)
V TACH	313	(102.18, 109.87)
VT $>$ 2	814	(91.88, 95.88)
V BRADY	17	(58.43, 74.27)
COUPLET	2833	(87.59, 89.26)
BIGEMINY	743	(79.69, 82.26)
PAUSE	655	(52.02, 55.46)
TACHY	10470	(115.11, 115.74)
PVC	1895	(89.61, 91.48)
ARTIFACT	39	(69.01, 76.06)
HR HI	8252	(123.29, 123.85)
HR LO	7556	(47.06, 47.32)
ARRHY SUSPEND	1300	(86.93, 88.98)

As is expected, many of the most important alarms correspond with low or high heart rate levels. HR Hi and HR Lo alarms are set to go off when the heart rate reaches unhealthy levels, thus the low variance in heart rate levels for these alarms is unsurprising. Other alarms, such as VFIB/VTAC, do not rely on heart rate levels, and thus heart rate data can be used to corroborate whether or not these alarms are true-positives and actionable.

For some patients, we also have blood pressure levels from the physiological data. This data is much more incomplete, but is useful in determining false positives vs true positives in some alarms.

3 Reducing Nuisance Parameter Alarms

There are several solutions for reducing the number and frequency of parameter alarms in the NICU. [5] offers several key insights, including allowing re-adjustment settings for parameter alarms. Using heart rate data, we note that re-adjustment settings may be necessary for heart rate alarms, where Hi HR and Lo HR alarms can compose over 30% of alarms for patients that have low or high mean heart rates.

The following two graphs show the proportion of HR Hi alarms vs patient mean HR and proportion of HR Lo alarms vs patient mean HR. We can see from Figure 1 and Figure 2 that the lack of re-adjustment settings can lead to HR parameter alarms dominating a patient's alarm data.

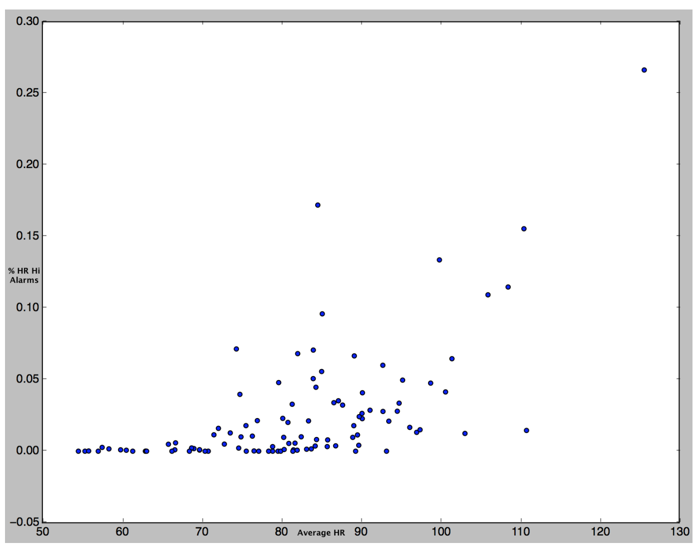

Figure 1: Example of Nuisance HR Hi alarms in some patients. The y-axis is % of HR Hi alarms, x-axis is patient's mean heart rates, n=102.

In a 2011 study, [6] suggested adding delays to reduce the number of nuisance parameter alarms. We used this strategy with the NICU data with several different delay intervals. We assume that a delay period begins once an alarm finishes. If an alarm of the same type begins during the delay period for the same patient, we remove this alarm. Figure 3 shows how we reduce the number

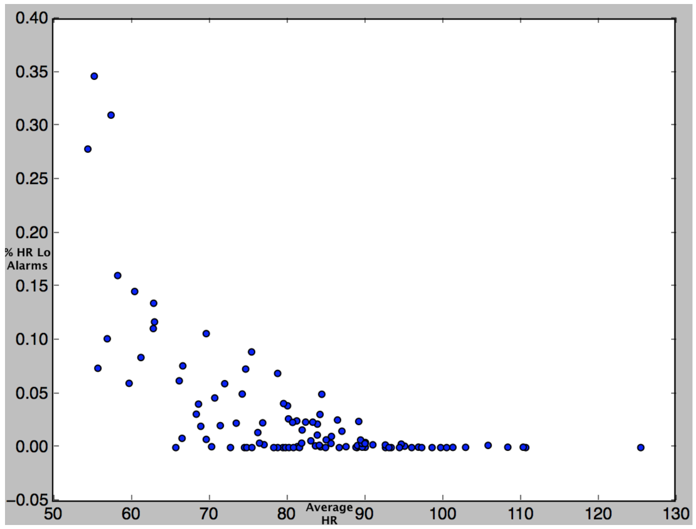

Figure 2: Example of Nuisance HR Lo alarms in some patients. The y-axis is % of HR Lo alarms, x-axis is patient's mean heart rates, n=102.

of parameter alarms with a variety of delay conditions. In fact, even a delay of ten seconds cuts down the number of parameter alarms by over 15%.

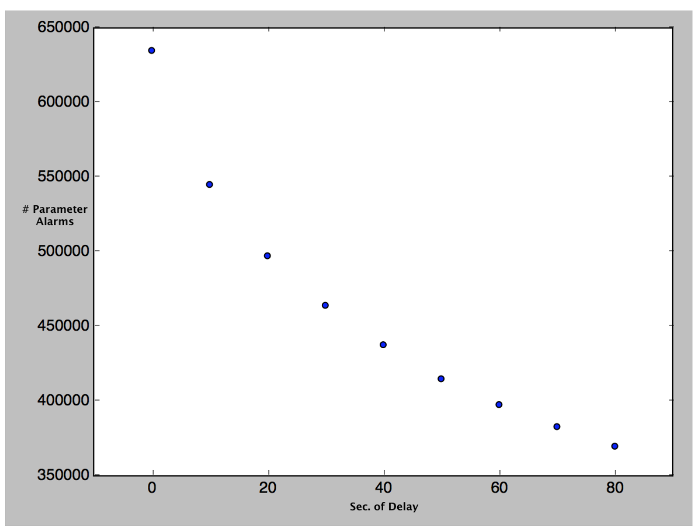

Figure 3: Example of delay decreasing the number of parameter alarms. The y-axis is the number of parameter alarms, x-axis is the number of seconds of delay.

4 Reducing Nuisance Crisis Alarms

4.1 Introduction

We believe that it is possible to use machine learning techniques to differentiate true positive crisis alarms and false positive crisis alarms. In order to do so, we must create a labeled data from alarm data and then train a classifier that can distinguish actionable from non-actionable crisis alarms. Defining features and labels for crisis alarms is an attainable task, but may require more data as well as medical knowledge.

4.2 Labeling

In [5], we get several standards for labeling true positives and false positives for a variety of alarms including ASYSTOLE, VFIB/VTAC, VTACH, PAUSE, and VBRADY.

- Examine electronic medical records (EMR) and confirm true positive alarms predicted patient complications.
- Confirm the correctness of crisis alarms by examining other leads, such as Pulmonary Artery pressure.

The physiological data occasionally includes the following three blood pressure measures: systolic pressure, diastolic pressure, and mean arterial pressure. Examining changes in blood pressure is helpful for teasing out true and false positives, but a medical professional should be consulted in order to confirm that the labels are correct.

Ultimately, because each alarm has different criteria for proving the presence of a true positive, it may be the case that each crisis alarm requires its own labeled data set and classifier. We will focus primarily on the ASYSTOLE alarm and the problem of distinguishing ASYSTOLE false and true positives. This is because an ASYSTOLE is classified as one of the most lethal crisis events, and thus an event in which false positives are particularly harmful.

The ASYSTOLE alarms sounds when the heart rate of a patient is at or near zero for 4-5 seconds. Unfortunately, sometimes these heart rate measures can be faulty. When blood pressure measures are present, false positives can be detected somewhat easily. Figure 4 bottom gives an example of a false positive. Notice that the heart rate dramatically drops with no change in blood pressure levels. Figure 4 top gives the example of a true positive. Notice that blood pressure levels reflect the change in pressure after the heart stops beating.

Unfortunately, while blood pressure levels can help distinguish true positives from false positives, the incomplete data means that most ASYSTOLE alarms are not accompanied by blood pressure levels.

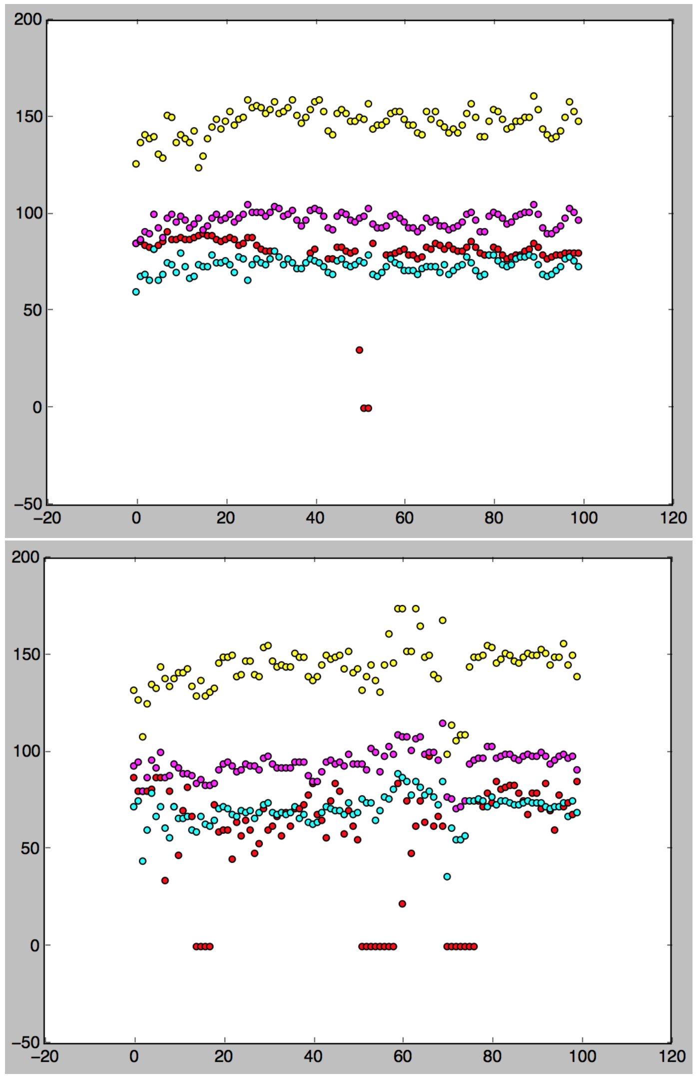

Figure 4: Top is an example of a false positive ASYSTOLE alarm, while bottom is an example of a true positive. The y-axis measures parameter levels and x-axis measures seconds. Heart rate measurements are plotted in red, systolic pressures are in yellow, diastolic pressures are in cyan, and mean arterial pressures are in magenta.

4.3 Feature Representations

There are a variety of features that we can use to predict whether ASYSTOLE alarms are actionable. Heart rate, blood pressure, and other physiological time-series data are immediately the best candidates. To turn time-series data into features, we can use the 60 second window prior to the alarm as the epoch length, and use the following statistical techniques as features (some drawn from [7]):

- RMSSD: Root mean square of successive difference between adjacent HRs in the window.
- SD: Standard deviation of HRs in the window
- MAX: Maximum HR in the window
- MIN: Minimum HR in the window
- MEAN: Average HR in the window
- MEDIAN: Median HR in the window

We can use the same statistics to add features for systolic blood pressure, diastolic blood pressure, and mean arterial pressure in the 60 second window. Other potential features could be patient specific, and perhaps take into account the risk score of a specific patient developing a life-threatening condition. In addition, previous alarm data and blood pressure levels can also be used as features (when present).

4.4 Techniques

Assuming we have a labeled data set, the next problem is to choose a Machine Learning technique that works well. This is essentially a two-class classification problem that can potentially be unbalanced depending on the proportion of false positives. There are several algorithms that excel at classification under these parameters:

- **Boosted Decision Trees**: Decision trees typically perform well on biased data sets. Boosted decision trees in particular can achieve high accuracy with relatively little training time.
- **SVM**: For many crisis alarms, the data set is only slightly unbalanced, making SVMs a feasible classifier. In addition, using SVMs in conjunction with under-sampling techniques or synthetic minority over-sampling techniques also could fix balancing issues.

4.5 Aside: Anomaly Detection and Smart Alarms

Another possible way to combat alarm fatigue is to create more robust, "smarter" alarms. Rather than respond to a certain threshold, these alarms could use machine learning techniques to analyze and detect anomalies in physiological data. Two techniques in particular are Symbolic Aggregate approXimation (SAX), which discretizes the input time series into a string, and Sequitur, which induces a context-free grammar (CFG) from it. These techniques in conjunction can be used to establish rules for heart-rate data and then find anomalies.

GrammarViz 3.0 is a program for time series exploration and analysis, that implements both SAX and Sequitur [8, 9]. We installed this software and were able to discretize the heart-rate levels of patient patient $id = 2222$. From this discretization, the program then attempts to detect heart-rate anomalies. Figure 5 shows the results of the anomaly detection algorithm.

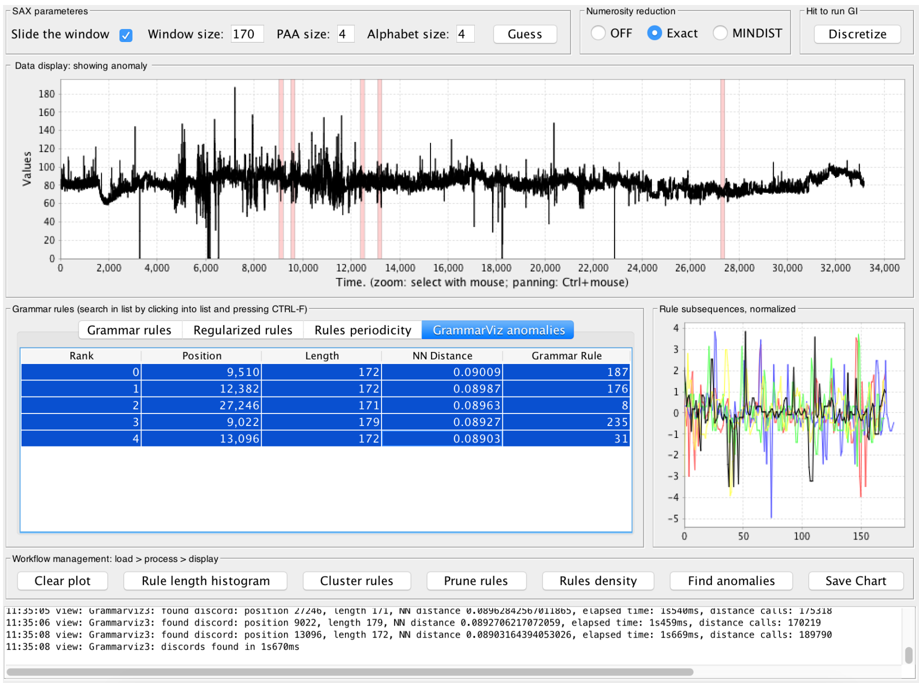

Figure 5: Results of GrammarViz 3.0 anomaly detection algorithm on patient $id = 2222$.

Without medical knowledge, it is difficult to determine the accuracy and feasibility of the anomaly detection algorithm. Ideally, one should be able to train the algorithm on a healthy heart-rate, and then use anomaly detection in real-time an unhealthy patient. GrammarViz 3.0 does not currently offer this possibility, but perhaps this could be a project for implementation in the future.

References

[1] Sue Sendelbach and Marjorie Funk. Alarm fatigue: a patient safety concern. *AACN Advanced Critical Care*, 24(4):378–386, 2013.

[2] L. Bell. Alarm fatigue linked to patient's death. interview by laura wallis. *The American journal of nursing*, 110(7):16–16, 2010.

[3] ACCI Healthcare Technology Foundation. 2016 national clinical alarms survey. 2016.

[4] ECRI Institute. 2017 top 10 technology hazards. *Health Devices.*, page 40, 2016.

[5] Barbara J. Drew et al. Insights into the problem of alarm fatigue with physiologic monitor devices: A comprehensive observational study of consecutive intensive care unit patients. *PLoS ONE*, 9(10), 2014.

[6] James Welch. An evidence-based approach to reduce nuisance alarms and alarm fatigue. *Biomedical Instrumentation and Technology: Alarm Systems.*, 45(s1):46–52, 2011.

[7] Stein P.K. Kleiger R.E. and Bigger J.T. Jr. Heart rate variability: measurement and clinical utility. *Ann Noninvasive Electrocardiol.*, 10(1):88–101, 2005.

[8] Pavel Senin, Jessica Lin, Xing Wang, Tim Oates, Sunil Gandhi, Arnold P Boedihardjo, Crystal Chen, Susan Frankenstein, and Manfred Lerner. Grammarviz 2.0: a tool for grammar-based pattern discovery in time series. In *Machine Learning and Knowledge Discovery in Databases*, pages 468–472. Springer, 2014.

[9] Pavel et al. Senin. Time series anomaly discovery with grammar-based compression. In *Proc. EDBT (Brussels, Belgium, March 2015)*, pages 481–492. 2015.