Machine learning techniques in intensive care monitoring

Keywords: intensive care, patient monitoring, alarm rules, false alarms

Wiebke Sieben, Karen Schettlinger, Silvia Kuhls
sieben, schettlinger, kuhls @statistik.tu-dortmund.de
Department of Statistics, Dortmund University of Technology, Germany

Michael Imhoff
mike@imhoff.de
Department for Medical Informatics, Biometrics and Epidemiology, Ruhr-University, Bochum, Germany

Ursula Gather
gather@statistik.tu-dortmund.de
Department of Statistics, Dortmund University of Technology, Germany

Abstract

Monitoring systems in intensive care units have a high false alarm rate. Machine learning techniques can be applied to improve existing alarm systems. We present two approaches, a filtering approach and a classification approach, and demonstrate their potential in reducing false alarms.

1. Introduction

Modern patient monitoring systems guarantee continuous surveillance of the patients' vital signs. Most alarm systems use thresholds for single vital signs which produce an alarm when the threshold for one of them is crossed. However, care givers often experience the alarms as nuisance. The frequency of false alarms has been subject to several studies which report false alarm rates of up to 90 % see for example Chambrin, Ravaux, Calvelo-Aros et al. (1999).

Transferring and verifying new alarm procedures for bedside patient monitoring in intensive care units is worthwhile. For this purpose, existing and feasible methods are investigated to reveal the "added value" of modern, complex, statistical modeling and machine learning methodology. In this context "added value" means primarily a reduced false alarm rate without a loss in sensitivity. Sensitivity is the probability that a system triggers an alarm in an alarm requiring situation. A false alarm is when a system triggers an alarm in a situation which does not require an alarm.

With an annotated reference data set of patient monitoring data at one-second intervals from a joint clinical study with the University Hospital Regensburg all methods may be tested under real world conditions. This data set contains vital signs, alarms and monitor settings of 85 cases, clinically evaluated by an intensivist. In approximately 1250 hours of observation 7000 alarms from an Infinity Monitoring System® (Dräger) have been annotated. Acquiring annotations is extremely time consuming and sets this data set apart from others.

2. Background

Methods and procedures for analyzing medical time series mostly aim at improvements in diagnosis, prognosis or therapy in medical applications. Augusto (2005) provides an overview of approaches concentrating on the time aspect in medical data analysis. Imhoff and Kuhls (2006) present a detailed summary of statistical approaches and alarm-algorithms for intensive care monitoring. Reviewing existing methods and procedures for intensive care monitoring, three main categories of approaches can be found:

(1) filtered time series + thresholds, e.g. (Koski et al., 1992)
(2) data driven alarm rule generation, e.g. (Tsien, 2000)
(3) knowledge based alarm rules, e.g. (Haimowitz et al., 1995).

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

3. Improving alarms in online monitoring

The first two categories address statistical techniques. In these categories we propose

(1) Repeated Median filtering for automatic real-time signal extraction + thresholds

(2) Neyman-Pearson modified Random Forests for data driven alarm rule generation.

3.1. Repeated Median filtering

Applying the alarm thresholds to the noise-free and artefact-free signal of a time series can yield a distinct reduction of the false alarm rate. The online Repeated Median filter (Davies et al., 2004; Gather et al., 2006) provides a useful tool for extracting signals without time delay from physiologic high-frequency time series which can contain measurement artefacts, level shifts, trends and sudden trend changes.

This filter applies the robust Repeated Median regression (Siegel, 1982) to a moving time window of fixed width n, i.e. a straight line is fitted to the last n observations and the signal is estimated online by the last fitted value within each time window. This filter is able to resist up to 50% artefacts within each time window and yields a smooth and reliable signal estimation which outperforms other well-known robust filters such as e.g. the running median especially if the observed time series contain a trend (Gather et al., 2006). Several sophistications of such a robust regression filter have been proposed including level shift detection, artefact removal and an adaptive choice of the window width. An overview over most of these methods is given in Schettlinger, Fried and Gather (2006).

The reference database was used to investigate the performance of threshold alarms based on the online Repeated Median signal. One third of the annotated data were used for tuning the new alarm algorithm, i.e. for determining the appropriate window size of the filter and for developing additional alarm rules.

For arterial blood pressure alarms the filter with a window width of 91 seconds performed best on the tuning dataset. When applying this alarm algorithm to the remaining data 45% of the alarms could be reduced in comparison to conventionally used threshold alarms for arterial pressures. At the same time a sensitivity of 90.7% for detecting alarm-relevant situations was obtained. One nice property of signal-based threshold alarms is shown in Fig. 2. In comparison to the simple threshold alarm system used in practice the new alarm-algorithm rather generates one long alarm for a certain situation instead of several short alarms.

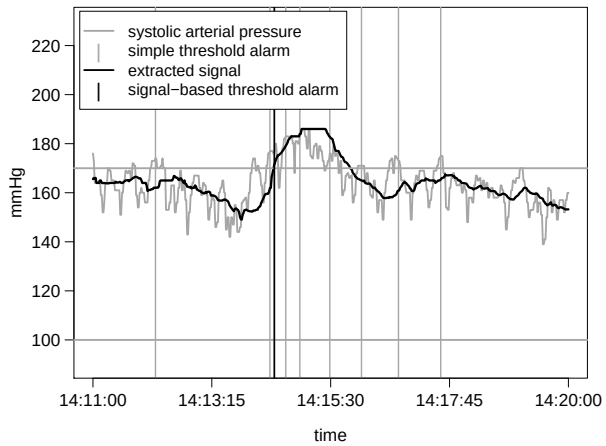

Figure 1. Example for signal-based threshold alarms in comparison to convetionally used threshold alarms

3.2. Neyman-Pearson modified Random Forests

Machine learning techniques, and particularly decision trees have proven suitable for alarm classification. Random Forests (Breiman, 2001), which are ensembles of randomized decision trees, can improve the accuracy compared to single trees in many situations. As the misclassification rate of non life-threatening situations is to be minimized under the constraint that the misclassification rate of life-threatening situations is close to zero, standard techniques like Random Forests need to be improved. In analogy to the Neyman-Pearson Lemma a classification rule is built as a hypothesis test for testing "situation is alarm relevant" vs. "situation is not alarm relevant" based on an ensemble of trees. This yields a classification rule for any given significance level which is the probability of misclassifying alarm-relevant situations (Sieben & Gather, 2007).

The performance of this procedure was analyzed on the reference data set. It was randomly divided into training and test sets 1000 times. On each such training set a modified Random Forest was grown containing 1000 trees. The forests were applied to the corresponding test sets and the reductions of false alarms and the obtained sensitivities were noted.

For a significance level of 5% the expected 95% sensitivity is well achieved on a test set with a mean sensitivity of 94,69% and a median sensitivity of 94,79% in 1000 generated modified Random Forests (Fig. 2). The variability in the achieved sensitivities is small:

no choice of training and test set results in a sensitivity below 90%. While keeping the sensitivity at about 95%, the false alarms are reduced by 51,27% (mean) and 51,53% (median).

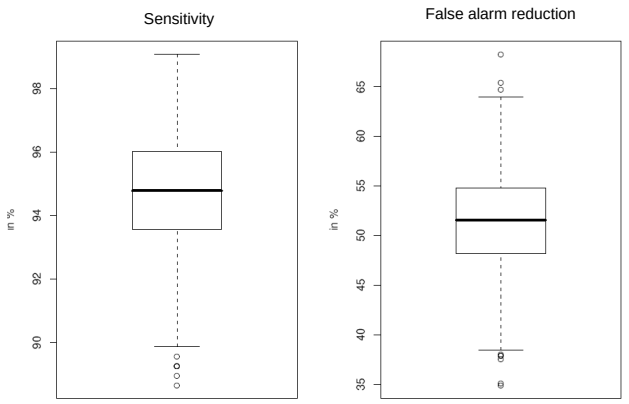

Figure 2. Sensitivities and false alarm reduction (significance evel 5%, full Bootstrap sampling)

From these results, we conclude that Neyman-Pearson modified Random Forests are suited for data driven alarm rule generation. However, repeating this evaluation with stratified training and test samples with respect to patients, such strong false alarm reductions cannot be achieved. Thus, the obtained alarm rules might not be easily transferred to other patient groups unless the data set is enlarged.

4. Conclusions

A time series filtering procedure and a classification approach were presented that aim at improving alarm systems in intensive care monitoring. Both the presented procedures are capable of reducing the false alarm rate with a high sensitivity. However, for practical applicability, the modified Random Forests need more training data. Robust Repeated Median filtering performs well for new patients. For this reason, this method could be implemented into a new monitoring system to lower the false alarm rate.

Acknowledgments

The support of the DFG (SFB 475) is gratefully acknowledged. The authors wish to thank Christian Wrede and Sylvia Siebig, Department of Internal Medicine I, University Hospital Regensburg, for their intense commitment to conducting the clinical study.

References

Augusto, J. (2005). Temporal reasinig for decision support in medicine. *Artificial Intelligence in Medicine*, *33*, 1–24.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Chambrin, M., Ravaux, P., Calvelo-Aros, D., Jaborska, A., Chopin, C., & Boniface, B. (1999). Multicentric study of monitoring alarms in the adult intensive care unit (icu): a descriptive analysis. *Intensive Care Medicine*, *25*, 1360–1366.

Davies, P., Fried, R., & Gather, U. (2004). Robust signal extraction for on-line monitoring data. *Journal of Statistical Planning and Inference*, *122*, 65–78.

Gather, U., Schettlinger, K., & Fried, R. (2006). Online signal extraction by robust linear regression. *Computational Statistics*, 33–51.

Haimowitz, I., Le, P., & Kohane, I. (1995). Clinical monitoring using regression-based trend templates. *Artificial Intelligence in Medicine*, *7*, 473–496.

Imhoff, M., & Kuhls, S. (2006). Alarm algorithms in critical care monitoring. *Anesthesia & Analgesia*, *102*, 1525–1537.

Koski, E., Mäkivirta, A., & Sukuvaara, T.and Kari, A. (1992). Developement of an expert system for haemodynamic monitoring: computerized symbolization of on-line monitoring data. *International Journal of Clinical Monitoring and Computing*, *8*, 289–293.

Schettlinger, K., Fried, R., & Gather, U. (2006). Robust filters for intensive care monitoring: Beyond the running median. *Biomedizinische Technik / Biomedical Engineering*, *51*, 49–56.

Sieben, W., & Gather, U. (2007). Classifying alarms in intensive care - analogy to hypothesis testing. *Proceedings of the Conference on 11th Conference on Artificial Intelligence in Medicine, Springer's LNCS* (pp. 130–138).

Siegel, A. (1982). Robust regression using repeated medians. *Biometrika*, *69*, 242–244.

Tsien, C. (2000). Event discovery in medical time-series data. *Proceedings of the AMIA Annual Fall Symposium 2000* (pp. 858–862).