Leveraging Routine Pre-Operative Blood Draws to Predict Hemorrhagic Shock During Surgery

Xinyu Li
Auton Lab
Carnegie Mellon University
Pittsburgh, PA 15213
xinyul2@andrew.cmu.edu

Michael R. Pinsky
School of Medicine
University of Pittsburgh
Pittsburgh, PA 15260
pinsky@pitt.edu

Gilles Clermont
School of Medicine
University of Pittsburgh
Pittsburgh, PA 15260
cler@pitt.edu

Artur Dubrawski
Auton Lab
Carnegie Mellon University
Pittsburgh, PA 15213
awd@cs.cmu.edu

Abstract

We investigate the utility of machine learning to estimate the likelihood of surgical patients to develop a hemorrhagic shock during surgery. We use hemodynamic responses to routine blood draws as virtual indicators of future shock. Our results suggest that by capturing sequential patterns in vital signs waveform data observed in pre-surgical settings, it is possible to predict which patients will likely get into shock in the course of surgery. These findings, after further validation, can enable building clinically useful screening tools to preempt complications in surgical patients and to inform medical resource allocation and planning, with the use of only already performed diagnostic procedures: routine lab test blood draws.

1 Introduction

Hemorrhagic shock, a medical condition caused by large amounts of lost blood, can induce multiple organ failures, produce adverse cellular effects, and even cause rapid death [2]. As certain types of surgeries may involve considerable blood loss [3, 4], determining the patients who are at high risk of hemorrhagic shock when subject to substantial bleeding could greatly help clinical assessment of patients' fitness pre-operatively and enable assignment of appropriate personnel and resources to ensure effective resuscitation if and when needed.

But, prediction of hemorrhagic shock from the hemodynamic waveform data collected in routine monitoring is a challenging problem due to the noisy nature of such data and diversity of patients' tolerances to acute blood loss. We hypothesize that predictability of shock can be sufficient for practical application if we consider responses to brief but rapid blood loss involved with blood draws taken routinely for laboratory tests just before surgeries. We bet on the assumption that responses to lab blood draws of patients who are at risk of developing hemorrhagic shock are distinct from those who are more robust to such complications.

In this paper, we describe a machine learning pipeline to identify the patients at high risk of hemorrhagic shock by leveraging sequential patterns that manifest in hemodynamic variables. We demonstrate that with the sequential patterns discovered in the waveform data collected during the routine pre-operative blood tests, our method is able to confidently identify 30% of the patients who are likely to experience hemorrhagic shock, outperforming logistic regression and random

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

forest models trained on statistically featurized data. In addition, the resulting compact graphical representations of discovered patterns allow clinicians to interpret the models and inferences easily.

2 Methods

We extracted sequential patterns from hemodynamic measurements using Graphs of Temporal Constraints (GTC), a method introduced by [1] to characterize and classify time-sequence datasets. A GTC represents a certain temporal pattern existing in a multivariate sequence of events. It has a form of a directed graph. Each vertex of a GTC represents a test of occurrence of an event of a particular type at a given time or position in a sequence, and it may be attached to any number of additional scalar tests, e.g. for the value of a variable against a threshold, at the same time or position. Edges between vertices of a GTC represent temporal constraints between tests. They are directed, which requires that the event at the destination vertex occurs within the specific time or position range of the event at the source vertex.

A GTC Decision Forest (GTC-DF) [1] is an ensemble model with components of a structure similar to standard decision trees, but using GTCs as nodes. Each such node in a tree partitions data into two complementary subsets depending on whether they contain the sequential pattern represented by the particular GTC or not. GTC-DF outputs a score for an unseen time sequence, which reflects the proportion of GTC trees classifying the time sequence to be positive. By training a GTC-DF model on the hemodynamic waveform data collected during routine test blood draws, we have identified sequential patterns across hemodynamic variables, that can serve as distinctive precursors of hemorrhagic shock to be developed during surgery, and we used them to make predictions about patients in a separate test cohort.

3 Experiments

3.1 Data

We have conducted our experiments on real-world laboratory animal data. n=36 healthy Yorkshire pigs were sedated and instrumented with arterial, pulmonary arterial and central venous catheters. They were stabilized for 30 minutes, and then bled at a constant rate of 20 mL/min. Hemodynamic variables, including arterial blood pressure, pulmonary pressure, central venous pressure, and airway pressure were collected at 250 Hz. 10 out of 36 pigs experienced hemorrhagic shock and received resuscitation as the bleeding proceeded. They were labeled as the positive class representing hemorrhagic shock outcome, and the other 26 pigs were labeled as the negative class.

Each subject had one blood test administered during the stabilization period before the start of controlled bleeding. A small amount of blood was drawn within a short time. The time stamp of blood draw was annotated by the technician who took it and recalibrated using pulmonary pressure artifacts associated with this test and present in data. We extracted the data of the three raw hemodynamic variables during the periods of blood draw unaffected by pulmonary pressure artifacts: Arterial Blood Pressure (ART), Central Venous Pressure (CVP), and Airway Pressure (AIR). Fig. 1 shows a small segment of data extracted for one subject.

3.2 Models

We concatenated extracted data from multiple subjects to train GTC-DF, as it does not require temporal alignment of time steps among subjects. The local maxima and minima (peaks) in the waveforms of the three hemodynamic variables were identified as characteristic events, and the original values of the variables as well as their basic statistics were used as scalar tests as defined in the GTC terminology. The peaks were found by detecting zeros of the first derivative of a variable after triangular moving average (TMA) smoothing was applied to the raw data. TMA was used to alleviate the high frequency noise and control the number of events marked.

The performance of GTC-DF was compared to logistic regression and random forest models on the binary classification task of predicting which subject is going to suffer from shock. Commonly used in hemodynamic monitoring moving statistical features were extracted to train logistic regression and random forest models, including sampled raw data points and means, standard deviations and

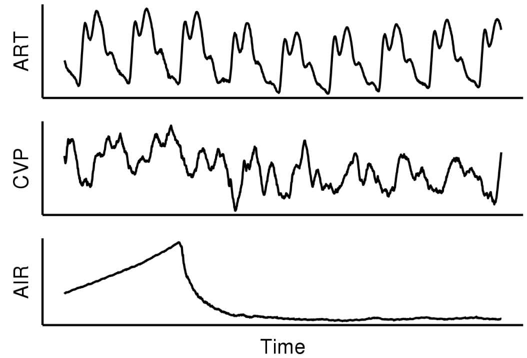


Figure 1: An example 5-second segment of three hemodynamic variables extracted from one subject.

Table 1: Model comparison

Model	TPR at low FPR (10^{-4})	TNR at low FNR (10^{-4})
Logistic regression	0.1	0.46
Random forest	0	0.23
GTC-DF	0.3	0.42

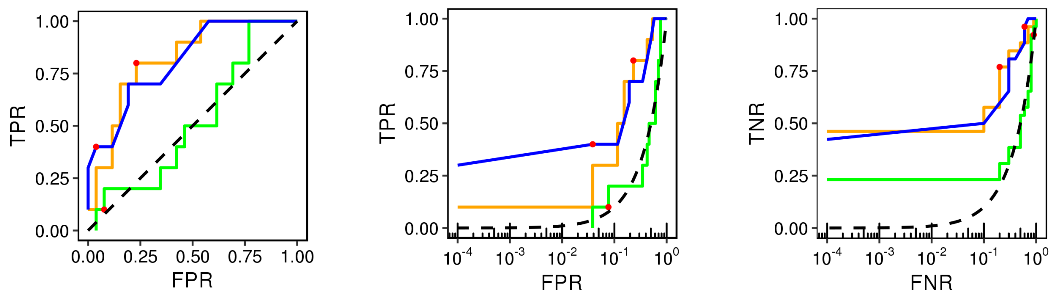
ranges, computed over a handful of time scales and updated at 1Hz. As the duration of blood draws varied among subjects, we used only the first 90 seconds of these periods for featurization, and then flattened the features so that each subject was represented with the same number of features.

To mitigate the risk of overfitting, we trained logistic regression with L2 regularization, and restricted the maximum depth of the trees in the random forest to 3. For GTC-DF, we executed an inner-loop leave-one-pig-out cross validation on each of the n=36 training sets in an outer leave-one-pig-out procedure, which was also used for logistic regression and random forest. Each inner loop cycle used data of (n-1)=35 subjects to build GTC decision tree, yielding 35 such trees to form a GTC-DF forest. We then used the median of the scores obtained from the n-1 GTC models as the overall prediction for the left-out 36th subject. We mapped the outputs of GTC-DF to probabilities by Platt scaling [5] and used them as the final predictions.

3.3 Results

Three models were evaluated in a leave-one-pig-out cross validation. In evaluation, we have focused at the corners of the Receiver Operating Characteristics (ROC) of the models, ie. in the ranges of sensitivity settings where the predictions can be made very highly confidently. These ranges of settings are the most important for clinical applications, in particular in quality of care improvement campaigns. We compared the performances of the models using the true positive rate (TPR) at low false positive rate (FPR) (10^{-4}) as the primary performance metric, assessing the ability to identify patients who are likely to develop hemorrhagic shock while having very low probability of giving false alarms. Correspondingly, we also report true negative rate (TNR) at very low false negative rate (FNR) to demonstrate the ability of the models to confidently identify patients who will likely not develop shock. These numeric results for each model are shown in Table 1, and the corresponding ROC curves in Fig. 2.

GTC-DF achieved the best performance among the three models we trained at the very low false positive rate setting: the recall rate of 0.3 at FPR 10^{-4}, and the 95%-ile confidence interval on recall of $[0.156, 0.444]$. This means that GTC-DF is able to confidently identify 30% (with the lower bound of 15.6%) the patients who are likely to develop hemorrhagic shock when exposed to substantial bleeding, when only giving one false alarm out of $10,000$ such predictions on average. Logistic regression achieved recall rate of 0.1 at FPR of 10^{-4}, with 95% confidence interval $[0, 0.201]$. GTC-DF was the only model with a lower bound of confidence greater than 0 at this setting, and conservatively, it could identify at least 15.6% of the positives. The logistic regression and GTC-DF



(a) (b) (c)

Figure 2: ROC curves of logistic regression (orange), random forest (green) and GTC-DF (blue). FPR and TPR corresponding to 50% sensitivity threshold for each model is marked with a red dot.

have shown a similar performance at the low FNR settings: both models were able to confidently identify about 45% of the negative patients. This capability is also very important for practical applications: clinicians want to know who will not require special anesthesiologist's attention during surgery, to inform optimal allocation of personnel and resources. We have found the random forest model to perform very well on training data, but overfitting heavily, even with depth of trees reduced to 3. It appears that diversity among subjects' responses to blood draws undermined its performance, while regularized logistic regression was still able to generalize reasonably well. We attribute the GTC-DF's robustness to overfitting when compared to random forest to its ability to extract sequential information from data. This can also explain GTC-DF's better recall at very low false detection rates when compared to logistic regression.

4 Conclusion

We used machine learning to leverage sequential patterns present in hemodynamic waveform data collected in clinical settings, and applied it to a task of a significant clinical importance. With the use of only a small amount of data collected during pre-operative routine blood tests, the proposed approach has been able to very confidently identify a substantial number of patients at risk for hemorrhagic shock. This has been possible using a relatively small cohort of n=36 subjects, 10 of whom belonged to the positive class. Our results suggest that by capturing sequential patterns in vital signs waveform data observed in pre-surgical settings, it is possible to predict which patients will likely get into shock in the course of surgery. It is also possible to identify with high confidence who will not likely get into shock. These findings can have profound implications on clinical practice by helping predictively manage surgical patients, and by informing allocation of appropriate personnel and resources on a case-by-case basis. Our future work includes validation of the presented protocol on a larger set of data, including data collected from human subjects. We also plan to better identify practical utility of the graphical models produced by GTC method in clinical interpretation of their meaning. We hope that the end result of our investigations will have a form of a clinically useful screening tool.

References

[1] M. Guillame-Bert and A. Dubrawski. Classification of time sequences using graphs of temporal constraints. *Journal of Machine Learning Research*, 18(121):1–34, 2017.

[2] G. Gutierrez, H. D. Reines, and M. E. Wulf-Gutierrez. Clinical review: hemorrhagic shock. *Critical Care*, 8(5):373, 2004.

[3] S. S. Hu. Blood loss in adult spinal surgery. *European Spine Journal*, 13(S01):S3–S5, June 2004.

[4] T. Mastracci, M. Bhandari, R. Mundi, S. Rizoli, B. Nascimento, and M. Schreiber. Operative blood loss, blood transfusion and 30-day mortality in older patients after major noncardiac surgery. *Canadian Journal of Surgery*, 55(6):426–428, dec 2012.

[5] J. C. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In *Advances in Large Margin Classifiers*, pages 61–74. MIT Press, 1999.