Predicting Mortality of Intensive Care Patients via Learning about Hazard

Dae Hyun Lee[1] and Eric Horvitz[2]

Biomedical and Health Informatics, University of Washington, Seattle, WA USA[1]; Microsoft Research, Redmond, WA USA[2]
dhlee4@uw.edu[1]; horvitz@microsoft.com[2]

Introduction

We explore the promise of employing machine learning and inference to assist with the care of patients in intensive care units (ICU). Such patients are acutely ill and have the highest mortality rates of hospitalized patients. Predictive models and planning systems could forecast and guide interventions to prevent the hazardous deterioration of patients' physiologies. Such analyses are supported by the relatively large amounts of data available about ICU patients, stemming from the use of specialized monitoring equipment and the intensive clinical workflow.

Identifying patients who are at high risk of death may be useful for guiding focus of attention on patients, including proactive monitoring and interventions. As the cost per day is typically higher in the ICU than for the general wards (Barrett et al. 2014), caregivers seek to move patients out of the ICU as soon as they believe that they are stable. Inferences about risk of death can guide decisions about retaining patients in the ICU or moving patients from general wards to the ICU.

Prior work on data-centric approaches has demonstrated boosts in performance for predicting the risk of death for ICU patients (Johnson et al. 2012) over the use of more widely used heuristic scores (Zimmerman et al. 2006). We report on the construction of a prediction pipeline that estimates the probability of death by inferring rates of hazard over time, based on patients' physiological measurements. The inferred models can provide insights to physicians about the contribution of each variable and information about the influence of sets of observations on the overall risks and expected trajectories of patients.

Data Description

We base our predictions on the PhysioNet ICU challenge dataset (Silva et al. 2012). This data was collected during the first 48 hours of observation after ICU admission for 4,000 patients, including clinical measurements, demographic information, and outcome-related information (i.e., in-hospital death, SAPS, length of stay). 13.85% of the patients died in hospital. For the study, we merged the non-invasive and invasive measurements for each blood pressure measurement (mean arterial pressure, systolic arterial blood pressure, diastolic arterial blood pressure) into single features. Fraction of inspired oxygen, *FiO2*, and partial pressure of oxygen in arterial blood, *PaO2*, were merged to constitute the *PaO_2/FiO_2 ratio (PFR)*, a widely used clinical variable to estimate pulmonary function. Since a clear indication of when patients are weaned from a ventilator machine could not be obtained from the data, mechanical ventilation, *MechVent,* was used as a demographic feature to present whether or not the patient had been mechanically ventilated within the first 48 hours following admission.

Method

With an eye to parsimony and associated interpretability of the model, we explore the power of employing a cumulative hazard function with a single governing parameter. We learn the parameter from data and then perform inference with the parametric function. A higher lambda drives the function to reach a higher probability of death earlier, so a learned lambda can be regarded as a quantitative summary of a patient's physiological stability.

We determine the probability of a patient death 48 hours after admission to the ICU. For each patient, time series data for each 48-hour period were divided into multiple time intervals $T_i, T_i = (t_i, t_{i+1}]\ where\ |T_i| = constant$, and the lambda for each interval λ_{T_i} was predicted using the trained model. Thereafter, the probability of death for each T_i was calculated using the cumulative hazard function of $p(\text{death}|T_i) = 1 - \exp\{-\lambda_{T_i} * (t_{i+1} - t_i)\}$. The probability of death at the 48[th] hour of ICU admission was then calculated with an absorbing Markov chain (Figure 1).

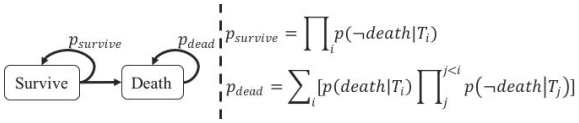

Figure 1. Absorbing Markov chain

Each patient-T_i instance is prepared in a way that summarizes the physiological condition of a patient during T_i. We calculated summarizing statistics (minimum, maximum, and average) as key observations within T_i. In addition, we performed statistical tests and resulting p values were used to quantify atypical values and the instability of variables (Table 1).

Feature	*Statistical Test*	*Alternative Hypothesis*
[Variable]_TT	*Two-tailed t-test*	$H_1: \mu_{training} \neq \mu_{sample}$
[Variable]_LT	*One-tailed t-test*	$H_1: \mu_{training} > \mu_{sample}$
[Variable]_GT	*One-tailed t-test*	$H_1: \mu_{training} < \mu_{sample}$
[Variable]_FT	*f-test*	$H_1: \sigma_{training} < \sigma_{sample}$

Table 1. Atypia sustainment and instability quantification

As we aim to predict the lambda based on the feature vector, we need to estimate the label for each training instance. The lambda for instances of deceased patients was calculated assuming the probability of death converged to 1 on the last date of ICU (LOS):

$$\lambda = -\ln[1 - \mathrm{p(death)}] / LOS,\ p(death) \approx 1$$

To estimate the lambda for instances of surviving patients, K-means clustering was conducted for all patient-T_i instances to group instances with similar physiology. Then, the lambda for each group was calculated based on given outcome variables as follows:

$$\lambda_{\mathrm{group}} = \sum Death / \sum Length\ of\ stay$$

The estimated lambda, λ_{group}, was assigned to each training instance of surviving patients within each cluster. Given the heterogeneity of ICU patients (i.e., different source of admission and different prognosis after ICU admission), we used an ensemble model. We used a regression random forest to train the model because the approach can present the distinct contributions of features to the model.

The hyper-parameters for the final model were selected from the model with best AUC from 10-fold cross-validation on training data. We performed evaluations on held-out test data using the predicted lambdas for each patient-T_i. The probability of death for these instances was computed using the predicted lambda. Thereafter, the resulting probabilities were aggregated using the absorbing Markov chain. For each validation trial, 90% of the patients were used as the training set, while 10% were allocated to the test set.

Results

T_i Width	Max score*	Threshold	AUC
8 hrs	0.4862±0.0587	0.4287±0.0587	0.8381±0.0267
12 hrs	0.4872±0.0340	0.4425±0.0340	0.8369±0.0231
24 hrs	0.4950±0.07103	0.4630±0.0368	0.8235±0.0294
48 hrs	0.4711±0.0548	0.4502±0.0469	0.8170±0.0269

*Table 2. Cross-validation results, K-means 20 group, *Score = min (Sensitivity, PPV)*

Discussion

The results (Table 2) show higher performance compared to the heuristic score (SAPS, 0.3125), with performance comparable to the top 10 scores reported in the PhysioNet competition (0.4513~0.5353). The maximum score was achieved in a 24-hour-window, which suggests a tradeoff between T_i width and model performance. Narrower T_i width might be able to detect local patterns while wider T_i width might be associated with fewer missing values.

Variables selected as most informative by the methodology overall aligned with observations employed in the heuristic score (See Table 3). However, we found that both *instability* and *sustainment quantifiers* were selected as top-ranked features. The latter results underscore the importance of capturing temporal dynamics for inferences about the risk of death from physiological measures.

Rank	8 hrs	12 hrs	24 hrs	48 hrs
1	Age	Age	Age	GCS_GT
2	GCS_MAX	GCS_MAX	SysABP_AVG	Urine_AVG
3	Urine_AVG	HR_MIN	GCS_MAX	SysABP_AVG
4	HR_MIN	Temp_FT	BUN_AVG	Age
5	HR_AVG	HR_AVG	HR_TT	HR_MIN

Table 3. Top 5 contributing features based on information gain

Conclusion

We constructed a predictive model for the risk of death of ICU patients via learning and inferring patient-specific mortality hazard rates. We found that the hazard model performed comparably with top data-centric models. The proposed method frames a promising direction of work on using an inferred hazard rate to characterize the trajectory of patients' deterioration over time and to build insights about the influence of specific observations on the risk of mortality.

References

Johnson, A.; Dunkley, N.; Mayaud L.; Tsanas, A.; Kramer, A. A.; and Clifford, G. D. 2012. Patient Specific Predictions in the Intensive Care Unit using a Bayesian Ensemble. *Computers in Cardiology* 39: 249-252.

Zimmerman, J. E.; Kramer, A. A.; McNair, D. S.; Malila, F. M.; et al. 2006. Acute physiology and chronic health evaluation (apache) iv: hospital mortality assessment for today's critically ill patients. *Critical care medicine* 34(5):1297– 1310.

Barrett, M. L.; Smith, M. W.; Elixhauser, A.; Honigman, L. S.; and Pines, J. M. December 2014. Utilization of Intensive Care Services, 2011. HCUP Statistical Brief #185. Agency for Healthcare Research and Quality, Rockville, MD.

Silva, I.; Moody, G.; Scott, D. J.; Celi, L. A.; and Mark, G. R. 2012. Predicting In-Hospital Mortality of Patients in ICU: The PhysioNet/Computing in Cardiology Challenge 2012. *Computers in Cardiology* 39: 245-248.