Prediction of Mortality in an Intensive Care Unit using Logistic Regression and a Hidden Markov Model

Srinivasan Vairavan[1], Larry Eshelman[1], Syed Haider[1], Abigail Flower[1], Adam Seiver[2]

[1]Philips Research North America (PRNA), Briarcliff Manor, NY, USA
[2]Therapeutic Care, Philips Healthcare, Carlsbad, CA, USA

Abstract

Intensive care medicine is a large share of the health care budget, and in the last decade there has been an increasing focus on making intensive care medicine more cost-effective by the efficient use of resources while still providing the best outcome for critically-ill patients. One important set of tools to perform this are critical illness severity assessment scores such as Simplified Acute Physiology score (SAPS-I) which help clinicians prioritize resources and determine the appropriate diagnostic/therapeutic plan for each patient. These scores are also used for assessing how medications, care guidelines, surgery, and other interventions impact mortality in Intensive Care Unit (ICU) patients. In an attempt to develop an improved patient-specific prediction of in-hospital mortality, we propose an algorithm based on logistic regression and Hidden-Markov model using vital signs (vitals), laboratory values (labs) and fluid measurements that are commonly available in ICUs. The algorithm was trained on 4000 ICU patient records and was validated on two sets of unseen test data of 4000 ICU patients each. These datasets were obtained as a part of PhysionNet/CinC Challenge 2012 (prediction of the mortality in ICU). Two different metrics, namely, (Event1) the minimum of sensitivity and positive predictive value and (Event2) a goodness of fit measure (range-normalized Hosmer-Lemeshow (H) statistic) was used to assess the algorithm's performance. The proposed algorithm achieved an Event 1 score of 0.50, 0.50 and an Event 2 score of 15.18, 78.9 compared to SAPS-I (Event 1: 0.3170, 0.312 and Event 2: 66.03, 68.58) in the two different validation dataset respectively. Furthermore, since the proposed algorithm uses instantaneous values of vitals and labs, it could be used as a continuous, real-time patient specific indicator of mortality risk.

1. Introduction

ICUs are responsible for an increasing percentage of the health care budget, and thus are a major target in the effort to limit health care costs [1]. Hence, there is an increasing need, given the resource availability limitations, to make sure that additional intensive care resources are allocated to those who are likely to benefit most from them. Critical decisions include interrupting life-support treatments and issuing do-not-resuscitate orders when intensive care is considered futile. In this context, mortality assessment is a crucial task, being used to predict not only the final clinical outcome but also to evaluate ICU effectiveness, and allocate resources.

Since the early 1980s clinical scores have been developed to assess severity of illness and organ dysfunction in the intensive care unit (ICU) setting [2, 3]. In the context of intensive medicine, severity scores are instruments that stratify patients according to a risk assessment based on clinical information about the patient. For example, SAPS-I is one such score which is widely used to account for population differences in studies aiming to compare how medications, care guidelines, surgery, and other interventions impact mortality in ICU patients [4]. Tools like SAPS-I have been used to improve the quality of intensive care and guide local planning of resources.

This paper presents the results of an improved patient-specific ICU mortality assessment algorithm. It makes use of several computationally sophisticated techniques.

One of the most promising recent innovations in the development of algorithms for use in ICUs is the use of data-mining and machine learning techniques [2]. The goal of data mining is to discover interesting patterns from raw data by using automatic discovery tools [5]. Machine learning techniques include such diverse techniques as rule-based approaches, artificial neural networks, logistic regression, and support vector machines.

Another set of techniques being applied to intensive care problems are dynamic statistical models such as the Hidden Markov Model (HMM) [6]. The underlying assumption of the statistical model is that the signal can be well characterized as a parametric random process, and the parameters of the stochastic process can be determined (estimated) in a precise, well-defined manner

[6]. Recently, state transition (Markov) models have been used to analyze progression of chronic diseases [7].

In an attempt to develop a patient-specific prediction of in-hospital mortality, we propose an algorithm based on logistic regression and HMM using vitals, labs and fluid measurements that are commonly available in ICU.

The data used to develop/validate the algorithm was obtained as part of PhysioNet/CinC Challenge 2012. The data consist of demographics, hourly measurements of vital signs and lab test results of 12,000 ICU patients. The proposed algorithm was trained on 4000 ICU patients (set A) (for which the gold standard of in-hospital mortality was provided) and it was validated on two different datasets of 4000 ICU patients each (set B and set C). The performance of the algorithm was assessed for its detection capability based on Event1 (minimum of Sensitivity and Positive Predictive Value) and prediction capability based on Event 2 (a range-normalized Hosmer-Lemeshow (H) statistic [8]). The result of the proposed algorithm was compared with SAPS-I, one of the commonly used critical illness severity assessment scores.

The paper is organized as follows: Section 2 presents the ICU clinical data, details of the proposed algorithm and its features. Section 3 describes the results and briefly discusses the rationale for the proposed algorithm. Finally, Section 4 makes some concluding remarks.

2. Methods

2.1. Intensive care unit data

The data used to develop the algorithm was obtained as part of PhysioNet/CinC Challenge 2012. The data consist of up to 41 different variables, which include general descriptors (age, gender, height, weight) and time series (hourly measurements of vital signs and lab test results) from the first 48 hours of the first available ICU stay of each of 4,000 patients chosen at random from a larger set. Each patient record has a retrospective gold standard of in-hospital mortality. Patients under the age of 16 and whose initial ICU stays were shorter than 48 hours (approximately the median) are excluded; there are no other exclusion criteria. For more details regarding the datasets, please refer to [9].

In addition to the 4000 ICU patient records, two unseen test datasets of 4000 ICU patient each are used for the validation of the proposed algorithm.

2.2. Proposed Algorithm

Two different algorithms were developed targeting a good performance in Event 1 and Event 2, respectively. The proposed algorithm is a combination of these two algorithms for their respective events. The details of each algorithm and its features are described below.

2.3. Event 1 specific

A general framework of the Event 1 specific algorithm is shown in Figure 1. The Event 1 specific algorithm uses logistic regression to combine different features such as age, PF ratio (PaO2/FiO2), creatinine, blood urea nitrogen (BUN), bilirubin, Glasgow Coma Score, glucose, disease development (sepsis and pneumonia), and the output of HMM. The different features used target oxygen transport and multi-organ system dysfunction for mortality risk assessment. The optimal threshold is chosen as the one that maximized the Event 1 metric in the training dataset of 4000 ICU patient records.

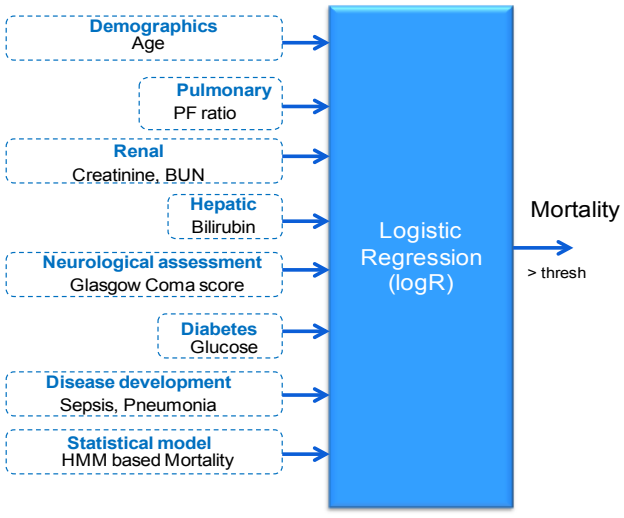

Figure 1. A general framework of Event 1 specific mortality prediction.

Hidden Markov Model (HMM) has been used for the last twenty years in the field speech recognition and in health care in both industry and academia; it is a dynamic model in nature that makes predictions over time [6]. The basic idea behind HMM is the estimation of hidden states of a process using observed variables over time. Figure 2 represents the general framework of the HMM based mortality prediction.

HMM requires three probabilities for each variable 'V', namely, the probability of each variable over the entire population data, P(V), the probability of each variable for only the people who lived, P(V|A), where 'A' is Alive, and finally the probability of each variable for the people who died, P(V|D), where 'D' is Dead. All the probabilities discussed above are calculated from the data.

The different possible state transitions are shown in Figure 3.

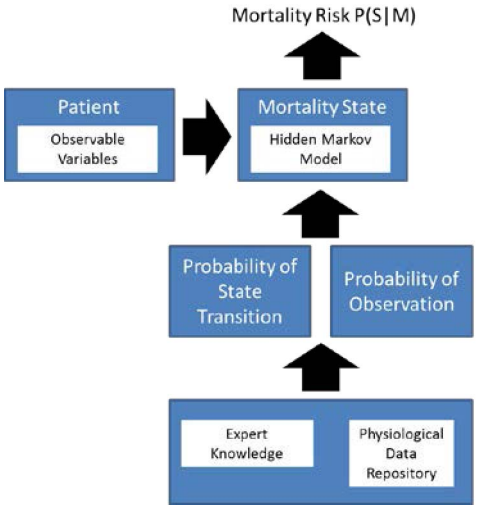

Figure 2. Estimation and prediction of critically ill patient's mortality state and the associated risk, P(S|M), the probability of state given measurements.

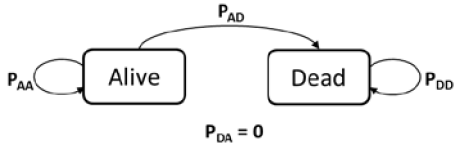

Figure 3. Markov chain for critically ill patient's mortality state.

The Markov Chain (MC) shown in Figure 3 provides more insight into critically ill patient's changing mortality state over time, using available measurements (vitals and labs) in the ICU.

To account for trend in patient's state, a mortality state sequence is computed. For example, if three measurements are available over the patient's ICU stay, the possible state transition sequences for the patient's mortality state are: $\{(A_1,A_2,A_3), (A_1,A_2,D_3), (A_1,D_2,D_3), (D_1,D_2,D_3)\}$. For m measurements, there are $m+1$ possible state sequences. HMM provides the state sequence that best fits the patient's mortality state.

The choice of variables used is based on the observations and expert knowledge of the ICU clinicians with many years of experience (Table 1). The variables are selected in such a way that when a critically ill patient's oxygen transport performs poorly, measurements at that point in time from the related variables raises the mortality risk. Subsequently when the patient goes into organ dysfunction, other than liver, the measurements from other related variables will further raise the mortality risk. Eventually when the patient goes into liver dysfunction, the risk computed at that point is likely the highest mortality risk.

The Event 1 score achieved by HMM is 0.49, and the Event 2 score is 183 (relatively high for Event 2) on training dataset. Therefore, HMM is combined with other features as shown in Figure 1 through a logistic regression model.

Table 1. Variables used for HMM based mortality detection.

Oxygen Transport	Organ Dysfunction other than Liver	Liver Dysfunction	Demographic
Lactate	Creatinine	Bilirubin	Age
PaCO2	WBC	ALT	ICU Type
SysABP (NI)	TroponinT	AST	
DiasABP (NI)	HCT		
HR	UrineDaily		
RespRate	Mg		
SaO2	Na		
PFratio	Glucose		
pH	GCS		

2.4. Event 2 specific

A general framework of the Event 2 specific algorithm is shown in Figure 4. The Event 2 specific algorithm is based on logistic regression and it utilizes three features, namely, Age, Urine Output/day and Glasgow Coma score for the prediction of mortality.

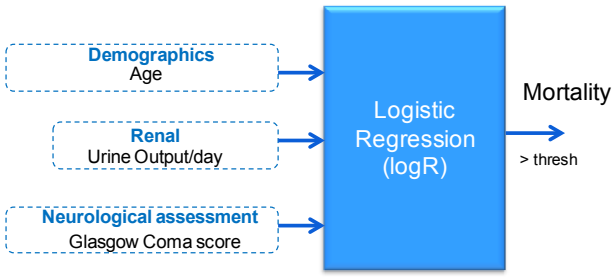

Figure 4. A general framework of Event 2 specific mortality prediction.

3. Results and discussion

The results of the Event 1 specific and Event 2 specific algorithms on training dataset (set A) and validation dataset (set B) are shown in Table 2. The proposed algorithm combines the two algorithms for their respective events. The SAPS-I scores for the training and validation dataset are also shown in Table 2. The proposed algorithm is then validated with an additional unseen test dataset (set C) and the results are shown in Table 3.

Table 2. Results on training (set A) and validation dataset (set B).

Algorithm	Datasets			
	Training (set A)		Validation (set B)	
	Event 1	Event 2	Event 1	Event 2
Event 1 specific	0.52	26.9	0.50	35.2
Event 2 specific	0.44	5.5	0.43	15.2
SAPS-I	0.30	68.4	0.31	66.0

Table 3. Results on validation datasets (set B and set C).

Algorithm	Validation datasets			
	Set B		Set C	
	Event 1	Event 2	Event 1	Event 2
SAPS-I	0.31	66.0	0.31	68.58
Proposed Algorithm	0.50	15.2	0.50	78.9

As seen from Table 2 and 3, the performance of the proposed approach is better than that of SAPS-I. In addition, the performance of the proposed approach is comparable to the validation dataset except for Event 2 metric in set C (Table 3).

Although both parts of the algorithm use logistic regression (which was chosen because it is more transparent than approaches such as support vector machines), the features used are quite different. We found that just the three listed features worked together to produce the best Event 2 score on dataset A and held up well on dataset B. However, it did not hold up so well on dataset C. Adding certain other features improved the Event 1 score, but lowered the Event 2 score on the training dataset (A). In particular, adding the HMM-based feature, while improving the overall performance on Event 1, always lowered the Event 2 score by a small margin. Nevertheless, this algorithm was fairly consistent across training (A) and validation datasets (B and C) on Event 1.

4. Conclusion

A patient-specific prediction algorithm of in-hospital mortality has been presented. The proposed algorithm is based on logistic regression and a Hidden-Markov model using data (vitals, labs and fluids) that are commonly available in ICU. It outperforms the commonly used critical illness severity assessment score, namely, SAPS-I. In addition, since it uses the instantaneous values of vitals and labs, it could be used to obtain patient specific trajectories of mortality risk assessment. The patient specific trajectories of mortality risk could be used by the clinicians to help them prioritize resources and determine the appropriate diagnostic/therapeutic plan for each patient.

References

[1] Halpern NA, Pastores SM. Critical care medicine in the United States 2000-2005: an analysis of bed numbers, occupancy rates, payer mix, and costs. Crit Care Med 2010;38(1):65-71.

[2] Rosenberg A. Recent innovations in intensive care unit risk-prediction models. Curr Opin Crit Care 2002;8:321-330.

[3] Silva A, Cortez P, Santos MF, Gomes L, Neves J. Rating organ failure via adverse events using data mining in the intensive care unit. Artificial Intelligence in Medicine 2008;43:179-193.

[4] Le Gall JR, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D, Mercier P, Thomas R, Villers D. A simplified acute physiology score for ICU patients. Crit Care Med 1984;12(11):975-977.

[5] Hand D, Mannila H, Smyth P. Principles of data mining. Cambridge, MA: MIT Press; 2001.

[6] Rabiner RL. A Tutorial on Hidden Markov Models and selected applications in speech recognition. Proceedings of the IEEE 1989; 77(2):257-286.

[7] Salazar JC, Schmitt FA, Yu L, Mendiondo MM, Kryscio RJ. Shared random effects analyses of multi-state Markov models: application to a longitudinal study of transitions to dementia. Stats Med 2007;26(3):568-580.

[8] Shah BV, Barnwell BG. Hosmer-Lemeshow goodness of fit test for Survey data. Joint Statistical Meetings - Section on Survey Research Methods 2003; 3778-3781.

[9] http://www.physionet.org/challenge/2012/

Address for correspondence.

Srinivasan Vairavan
Philips Research North America (PRNA)
345 Scarborough Road, Briarcliff Manor, NY 10501, USA
Tel/Fax: +19149456006/6580
srinivasan.vairavan@philips.com