Mortality Prediction in ICUs Using A Novel Time-Slicing Cox Regression Method

Yuan Wang, PhD[1], Wenlin Chen[1], Kevin Heard[2], Marin H. Kollef, MD[2], Thomas C. Bailey, MD[2], Zhicheng Cui[1], Yujie He[1], Chenyang Lu, PhD[1], Yixin Chen, PhD[1]
[1]Department of Computer Science and Engineering, Washington University, St. Louis, MO; [2]School of Medicine, Washington University, St. Louis, MO

Abstract

Over the last few decades, machine learning and data mining have been increasingly used for clinical prediction in ICUs. However, there is still a huge gap in making full use of the time-series data generated from ICUs. Aiming at filling this gap, we propose a novel approach entitled Time Slicing Cox regression (TS-Cox), which extends the classical Cox regression into a classification method on multi-dimensional time-series. Unlike traditional classifiers such as logistic regression and support vector machines, our model not only incorporates the discriminative features derived from the time-series, but also naturally exploits the temporal orders of these features based on a Cox-like function. Empirical evaluation on MIMIC-II database demonstrates the efficacy of the TS-Cox model. Our TS-Cox model outperforms all other baseline models by a good margin in terms of AUC_PR, sensitivity and PPV, which indicates that TS-Cox may be a promising tool for mortality prediction in ICUs.

Introduction

Over the last few decades, machine learning and data mining techniques are increasingly used for ICU mortality prediction. For example, PhysioNet/CinC organized a challenge in 2012 to develop methods for patient-specific prediction of in-hospital mortality [1], where the data includes time series of vital signs during the 48 hours after ICU admission. There are two tracks in this competition concerning binary outcome measurement (dead or alive) evaluated by sensitivity and positive predictive value (PPV), and risk estimation evaluated by a range-normalized Hosmer-Lemeshow statistic. A great deal of models submitted by participants largely outperformed the baseline algorithm (SAPS-1). For example, Johnson *et al* [2] used a novel Bayesian ensemble learning algorithm, and Krajnak *et al* [3] combined machine learning and clinical rules to build a predictive model.

Existing algorithms do not fully take advantage of the temporal changes of the input clinical time-series data. Models such as logistic regression (LR), support vector machine (SVM), and decision trees (DT) can only handle static and fixed-length feature vectors, which greatly limit their use on time-series. In this regard, it makes more sense to resort to Cox regression since it takes into account the temporal orderings of various features and it has widespread uses in survival analysis. In the following, we mainly survey the literature related to Cox regression in the area of medical care utilization.

Cox regression [4,5], though invented in the seventies, is still receiving significant interests from the domains of both healthcare and information technology. Different from methods mentioned above, Cox regression is one of the few models that take into account the temporal structure of events, which is very important for using time-series data to predict morbidity, readmission and mortality and to identify the associated factors. One of the classical examples for predicting morbidity is the risk assessment of cardiovascular disease [4]. A few studies also applied the Cox model for the prediction of outcome in patients with diabetes [6,7], stroke prediction [8], as well as ecologic studies [9]. These algorithms use the vast domain-specific knowledge that has been accumulated on the disease to manually select a limited number of risk factors and then put them into a Cox model.

Recent studies have also recommended using various machine-learning techniques combined with Cox regression for diverse topics in healthcare [10,11]. In addition, there have been a great effort towards making Cox regression more generalizable by adding regularizations to the parameters [12-15].

The above methods typically work on relational data with factors such as age, gender, race, weight, ecologic exposures, lab variables and clinical risk factors. Although some factors may be time-varying, they are typically processed in the canonical way. However, few have used Cox regression to classify multi-dimensional time-series

data, due to the challenge that Cox regression is originally proposed for survival analysis and cannot be directly used in time-series classification.

To address this problem, we propose in this paper a novel approach named Time Slicing Cox regression (TS-Cox), which not only takes advantage of Cox regression, but also can be applied on clinical time-series data. Specifically, the rich time-series data allow us to used sophisticated features such as statistical moments, powers in frequency domain, estimates in chaos theory and entropies in information theory as a novel set of covariates to predict mortality risk. Then, based on time-to-event characteristics of ICU data, we divide each time series into multiple windows for each waveform and extract features from every sliding window via signal analysis. Next, we employ Cox regression on those windows together with their associated time stamp information. After the model is trained, we are able to compute a hazard ratio as risk score for each waveform based on the extracted features and learnt model. Our TS-Cox model allows us to leverage on both powerful time-series features and the survival information in the time domain. To our knowledge, this is the first work that adapts Cox regression to classify multi-dimensional time-series.

The remainder of this paper is organized as follows. First, we introduce some background on the original Cox model and describe the proposed TS-Cox model. Next, empirical results of our TS-Cox model compared with baseline methods are presented and discussed. Finally, we summarize the research work and draw conclusions.

Methods

Feature extraction method

The input of our model is based on the features extracted from the multiple time-series of each patient. Many of our features are inspired by classic signal analysis methods including RR series distribution patterns (Mean, Median, Variance, Skewness, Kurtosis), magnitude of variability in the time domain (SD, RMSSD), and linear estimates in the frequency domain (power of VLF, LF, HF and LF/HF). A complete recipe for extracting these physiological features can be found in [16].

Besides, we also make use of features kindled by chaos theory and information theory, which are helpful in analyzing the degree of self-affinity and randomness of the time-series. Specifically, we adopt the detrended fluctuation analysis (DFA) based on chaos theory to estimate the statistical self-similarity of a signal [17-19]. We also compute the randomness measurements including approximate entropy [20,21], sample entropy [22,23] and permutation entropy [24]. They are measurements designed to quantify the degree of regularity versus unpredictability, reflecting the unpredictability of fluctuation in a time-series. A low value of the entropy indicates that the time-series is deterministic while a high value means that the time-series is unpredictable. These entropies are good indicators for cardiovascular signals where the occurrence of disease is highly correlated with the decrease of entropy. In summary, in this study we extracted 16 features from two types of time-series: heart rate (HR) and peripheral capillary oxygen saturation (SpO2), leading to 32 features in total for each time window of a patient.

Classification method

Cox Regression

Cox regression is a popular statistical model in survival analysis which deals with analysis of time duration until the happening of one or more events, such as sudden deterioration of a patient [4]. It estimates the risk of the occurrence of a particular event at a particular time stamp by quantifying the relation between the features and the risks.

Now we describe the Cox regression model. Let $X=(x_1, x_2, \ldots, x_m)$ be the feature vector for each patient where m is the number of features. In this paper, the features come from the aforementioned signal analysis method. Let $h(t)$ be a hazard function assessing the instantaneous risk of an event (*e.g.* mortality) at time t. The Cox regression expresses the logarithm of the mortality risk $h(t)$ as a weighted sum of all features with a base function as follows:

$$\log h(t \mid X) = \alpha(t) + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m , \quad (1)$$

where β_i is the coefficient of feature x_i reflecting the relation between this feature and the risk. For example, a large β_i means that the raise of the value of feature x_i increases the risk of mortality. $\alpha(t)$ is a base function independent

from the feature vector X whose output varies over time depending on t. Given the above equation (1), we have that the mortality risk $h(t)$ can be expressed as follows:

$$h(t \mid X) = \underbrace{h_0(t)}_{\text{baseline hazard function}} \times \underbrace{\exp(\beta_1 x_1) \times \ldots \times \exp(\beta_m x_m)}_{\text{proportional hazards}} \quad (2)$$

where $h_0(t)=exp(\alpha(t))$ is called baseline hazard function and the following product is called proportional hazards. The proportional hazards assume that each feature contributes an independent multiplicative factor to the primary risk hazard rate. Cox regression estimates the parameters β using maximum likelihood estimation over the partial log-likelihood function in which $h_0(t)$ is cancelled out. Consequently, we do not need to know the functional form of $h_0(t)$ as the output does not depend on the feature vector. Given a fixed time stamp t, the risk hazard rate $h(t|X)$ only depends on the feature values. Therefore, when computing the risk hazard rate at a particular time stamp in the future, we only need to compute the proportional hazards as a proxy for risk hazard rate.

The problem of survival analysis is further compounded by the presence of censored data [25,26], where the status of an instance is not observable after a certain time point. For example, an instance in mortality analysis is censored when the patient at question is still alive at the end of observation. Censored times C_i are associated with each instance i along with observed time for the event O_i. Define the failure time T_i for instance i as the minimum of O_i and C_i, *i.e.* $T_i = \min(O_i, C_i)$. $O_i \leqslant C_i$ indicates that the event of interest has occurred within the censoring time. However, if O_i is unknown then T_i is set to C_i and the instance is censored. Take hospital readmission prediction as an example. An event is defined as the onset of heart failure readmission within 30 days of discharge from the previous admission. The censored cases can be identified when (i) the patients whose follow-up details were lost over time or (ii) the patients was not readmitted within the time period of follow-up until the end of the study (which is fixed to 30 days in this case). This is commonly called the right censoring setting, which is the most frequently studied censoring phenomena in survival analysis.

Time Slicing Cox regression (TS-Cox)

Now we introduce our Time Slicing Cox regression (TS-Cox) method, which applies the Cox model to time-series data. Our method contains three steps: preprocessing, training, and deployment. We describe these three steps in detail as follows.

Preprocessing. We assume that the time-series has a basic *unit step* which is the time gap between two measurements. For example, in our data from MIMIC-II, each unit step is 60 seconds. For each patient, we maintain a sliding window from the beginning of each of his or her time-series. The window is a fixed-length time frame (*e.g.,* window size =1024 unit steps) that can be regarded as a snapshot of the status of a patient. At each iteration, we extract features from the time window and record the time stamp of this window. We then move the window by a fixed step-size and repeat until we reach the end of the patient's time-series.

For each time window, we adopt the aforementioned feature extraction method to obtain features from the time-series within the window. In our case, each time window is associated with two time-series for HR and SpO2. We extract 16 features from each time-series separately, leading to 32 features in total for each time window.

In each iteration, we move the window by a pre-specified step size (*e.g.,*100 unit steps) and extract features again for the new window. This way, we have multiple windows for each patient. Suppose the number of windows for a patient is L. Each window is also associated with a new failure time, which is the time gap between the end of the window and the original failure time T. Let D_l be the last time stamp of window l, $l=1,\ldots,L$, and S_l be the failure time for window l, we have that $S_l=\max(T-D_l, 0)$. At the end of preprocessing, for each patient we have L windows where each window l is associated with 32 features and a failure time S_l reflecting the new survival time for that window.

Training. The training procedure is to estimate the value of β, which largely follows the same workflow for Cox regression. Different from the original Cox regression where each training instance is a patient, our TS-Cox model counts each window of each patient as a training instance. In other words, each window is viewed as a unique patient. This way, the TS-Cox model considers the time-varying migration of features and the temporal ordering of all the time windows. In contrast, if we extract features directly from the entire time-series for all patients and then input them into models such as LR or SVM, we would lose such temporal information.

We still learn a weight vector β using maximum likelihood estimation after training the TS-Cox model. For example, if there are 32 features for each time window, the length of the vector β is 32. The training process is to tune the value of β so that the resulting risk hazard rates for all training instances best respect the order of their failure time S_l.

Deployment. With the learnt β in place after training, the evaluation procedure is to evaluate the risk of mortality for an out-of-sample patient. Specifically, given a patient, we extract the 32 features from their last time window using the same window size as in preprocessing and training. We then compute the proportional hazard *h'(X)* which is the exponential of the dot product of β and the feature vector X, *i.e.* $h'(X) = \exp(\beta_1 x_1 + \beta_2 x_2 + ... + \beta_m x_m)$ as the hazard risk rate. To use the proportional hazard rate for binary classification, we need to set a threshold c so that those patients with $h'(X)>c$ belong to one class and the rest belongs to another class. The threshold c is typically chosen by aiming at a prescribed specificity, sensitivity, or positive predictive value (PPV) based on cross-validation on the training data.

In practice, our model can be deployed to monitor patients in real-time. We can move the time window to the last point as new data come in, extract updated features, and re-compute the hazard rate by feeding the updated features into the TS-Cox model. An alert will be triggered when the criterion (such as $h'(X)>c$) for certain event is met.

Results and Discussion

We conduct experiments to evaluate the performance of the proposed method. In our experiments, we used a publicly available ICU database, namely Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC II) [27]. The data harnessed in MIMIC-II were collected from the ICUs of Beth Israel Deaconess Medical Center from 2001 to 2008 and covers 26,870 adult hospital admissions (version 2.6). Two types of data were obtained: clinical data and physiological waveforms. The clinical data were acquired from the CareVue Clinical Information System (models M2331A and M1215A; Philips Healthcare, Andover, MA) and the hospital's electronic archives. The data included patient demographics, nursing notes, discharge summaries, continuous intravenous drip medications, laboratory test results, and nurse-verified hourly vital signs, etc. The physiological waveforms were collected from bedside monitors (Component Monitoring System Intellivue MP-70; Philips Healthcare) and included high-resolution (125 Hz) waveforms (*e.g.*, electrocardiograms), derived time series such as HR, blood pressures, SpO2, and monitor-generated alarms [28]. Here, we only extracted the time series of HR and SpO2 as the inputs to our model. Finally, our experiment datasets consist of 930 waveforms including complete HR and SpO2 records, among which there are 56 dead and 874 censored records. Our task is to predict whether a patient will die in the ICU.

We compare our method with the following baselines: 1) Classical classifiers including logistic regression (LR), linear SVM (SVM-l) and SVM with RBF kernel (SVM-r). Specially, we follow the traditional approach of first extracting features from the whole time series and then apply those classifiers for binary classification. 2) Time-Slicing version of logistic regression (TS-LR), linear SVM (TS-SVM-l) and SVM with RBF kernel (TS-SVM-r). In particular, we follow the same time-slicing mechanism as in TS-Cox and train these models on all time windows to learn the weight of each feature. When making prediction, we apply these learned weights on the last window of each waveform. For all time-slicing algorithms, we use 1024 unit steps as window size and 100 unit steps as step size for moving windows based on the previous experiments.

The dataset is highly imbalanced with the majority of patients being non-dead. Therefore, accuracy performance does not make the most sense in this case since one can achieve a high accuracy by always predicting non-dead for every patient. As a result, for evaluating the performance, we fixed the specificity of each model to about 0.95 and compare the sensitivity and PPV of each model. What's more, we also evaluate the area under the ROC curve (AUC_ROC) and area under the precision-recall curve (AUC_PR). All the results reported in Table 1 are based on 5-fold cross validation. That is to say, we divide the whole dataset into 5 subsets where each subset contains 186 waveforms. Each algorithm is evaluated on each subset with all other 4 subsets being the training dataset. The final result is the mean of all evaluated subsets.

Table 1 below shows the results of all algorithms with specificity being fixed to 0.95. We make the following observations:

First of all, the proposed Cox regression with time slicing is better than the original Cox regression regardless of any measurements, which demonstrates the efficacy of using time slicing techniques in the Cox regression model. The advantage of Cox regression is that it takes time-to-event into consideration, while other models do not. Therefore, time slicing techniques work for Cox regression, but do not help other models since it merely increases the training instances for these classifiers.

Second, TS-Cox has both the largest AUC_ROC and the largest AUC_PR. Furthermore, TS-Cox surpasses all other models by a larger margin in regard of AUC_PR than of AUC_ROC, which signify that it makes more sense to measure AUC_PR than to estimate AUC_ROC for a skewed class distribution (most people still alive).

Third, the TS-Cox model constantly outperforms all the baseline models in terms of sensitivity and PPV, which indicate that TS-Cox is superior to other prediction models. We attribute this promising progress to the fact that the TS-Cox model is able to exploit the rich time-variant information of clinical time-series data.

Therefore, TS-Cox has a potential advantage over other methods to predict mortality in adult ICU monitoring system.

Table 1. Statistical Performance

	AUC_ROC	AUC_PR	Specificity	Sensitivity	PPV
LR	0.7463	0.1642	0.9485	0.1352	0.1455
SVM-l	0.4669	0.0763	0.9496	0.0643	0.0764
SVM-r	0.6823	0.1247	0.9485	0.1012	0.1100
Cox	0.5783	0.1195	0.9575	0.1945	0.2267
TS-LR (Step=100)	0.7181	0.1826	0.9498	0.1652	0.2253
TS-SVM-l (Step=100)	0.5452	0.0938	0.9500	0.0456	0.0785
TS-SVM-r (Step=100)	0.6841	0.1402	0.9504	0.1336	0.1649
TS-Cox (Step=100)	**0.7514**	**0.2522**	0.9485	**0.3424**	**0.2956**

Conclusions

This paper proposed an integrated architecture for predicting death events in adult ICU patients based on a combination of time-series feature extraction methods and a novel time-slicing Cox regression method. The former produces more discerning features, and the latter exploits the strategy of slicing time windows on the static Cox model according to its time-to-event characteristics. We compared our TS-Cox model to conventional Cox model as well as LR, SVM-l, SVM-r, and their time-slicing version. The experiments indicate that TS-Cox has the highest AUC_PR, sensitivity and PPV by a huge margin at a specificity value around 0.95. These promising results demonstrate the feasibility of our new framework for better mortality prediction in adult ICU monitoring system.

Acknowledgements

This work was partially supported by CCF-1215302, IIS-1343896, and DBI-1356669 grants from the National Science Foundation of the USA, a Microsoft Research New Faculty Fellowship, a grant from the Barnes-Jewish Hospital Foundation, and a URSA grant from the Washington University.

References

1. Silva I, Moody G, Scott DJ, Celi LA, Mark RG. Predicting in-hospital mortality of ICU patients: The PhysioNet/Computing in Cardiology Challenge 2012. 2012. pp. 245–8.

2. Johnson AEW, Dunkley N, Mayaud L, Tsanas A, Kramer AA, Clifford GD. Patient specific predictions in the intensive care unit using a Bayesian ensemble. Computing in Cardiology (CinC), 2012. pp. 249–52.

3. Krajnak M, Xue J, Kaiser W, Balloni W. Combining machine learning and clinical rules to build an algorithm for predicting ICU mortality risk. Computing in Cardiology (CinC) 2012. pp. 401–4.

4. Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society Series B Methodological. 1972 Jan;34(2):187–220.

5. Cox DR. Partial likelihood. Biometrika. 1975 Aug;62(2):269–76.

6. Andersen PK, Borch-Johnsen K, Deckert T, Green A, Hougaard P, Keiding N, et al. A cox regression model for the relative mortality and its application to diabetes mellitus survival data. Biometrics, 1985 Dec;41(4):921–32.

7. Perkins BA, Ficociello LH, Silva KH, Finkelstein DM, Warram JH, Krolewski AS. Regression of microalbuminuria in type 1 diabetes. New England Journal of Medicine. 2003;348(23):2285–93.

8. Lumley T, Kronmal RA, Cushman M, Manolio TA, Goldstein S. A stroke prediction score in the elderly: validation and web-based application. J Clin Epidemiol. 2002 Feb;55(2):129–36.

9. Lepeule J, Rondeau V, Filleul L, Dartigues J-F. Survival analysis to estimate association between short-term mortality and air pollution. Environ Health Perspect. 2006 Feb;114(2):242–7.

10. Neuvirth H, Ozery-Flato M, Hu J, Laserson J, Kohn MS, Ebadollahi S, et al. Toward personalized care management of patients at risk: the diabetes case study. San Diego, California, USA: ACM KDD; 2011. pp. 395–403.

11. Yu S, van Esbroeck A, Farooq F, Fung G, Anand V, Krishnapuram B. Predicting readmission risk with institution specific prediction models. IEEE ICHI; 2013. pp. 415–20.

12. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med. 1997 Feb 28;16(4):385–95.

13. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. Journal of statistical software. 2010;33(1):1–22.

14. Vinzamuri B, Reddy CK. Cox regression with correlation based regularization for electronic health records. IEEE ICDM; 2013. pp. 757–66.

15. Vinzamuri B, Li Y, Reddy CK. Active learning based survival regression for censored data. Shanghai, China: ACM CIKM; 2014. pp. 241–50.

16. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart Rate Variability: Standards of measurement, physiological interpretation, and clinical use. European Heart Journal, 1996 Mar 1;93(5):1043–65.

17. Peng CK, Havlin S, Stanley HE, Goldberger AL. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. Chaos. AIP Publishing; 1995;5(1):82–7.

18. Yum MK, Park EY, Kim CR, Hwang JH. Alterations in irregular and fractal heart rate behavior in growth restricted fetuses. Eur J Obstet Gynecol Reprod Biol. 2001 Jan;94(1):51–8.

19. Nakamura T, Horio H, Miyashita S, Chiba Y, Sato S. Identification of development and autonomic nerve activity from heart rate variability in preterm infants. BioSystems. 2005 Jan;79(1-3):117–24.

20. Pincus SM. Approximate entropy as a measure of system complexity. PNAS. National Acad Sciences; 1991 Mar 15;88(6):2297–301.

21. Ho KK, Moody GB, Peng CK, Mietus JE, Larson MG, Levy D, *et al*. Predicting survival in heart failure case and control subjects by use of fully automated methods for deriving nonlinear and conventional indices of heart rate dynamics. Circulation. 1997 Aug 5;96(3):842–8.

22. Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol. 2000 Jun;278(6):H2039–49.

23. Lake DE, Richman JS, Griffin MP, Moorman JR. Sample entropy analysis of neonatal heart rate variability. Am J Physiol Regul Integr Comp Physiol. 2002 Sep;283(3):R789–97.

24. Bandt C, Keller G, Pompe B. Entropy of interval maps via permutations. Nonlinearity. IOP Publishing; 2002 Sep 1;15(5):1595–602.

25. Klein JP, Moeschberger ML. Survival analysis: statistical methods for censored and truncated data. Springer; 2003.

26. David W Hosmer J, Lemeshow S, May S. Applied survival analysis. Hoboken, NJ, USA: John Wiley & Sons; 2011. 1 p.

27. Saeed M, Villarroel M, Reisner AT, Clifford G, Lehman L-W, Moody G, *et al*. Multiparameter Intelligent Monitoring in Intensive Care II: a public-access intensive care unit database. Crit Care Med. 2011 May;39(5):952–60.

28. Clifford GD, Scott DJ, Villarroel M. User guide and documentation for the MIMIC II database. 2012 Feb.