Artificial Intelligence in Medicine 63 (2015) 191–207

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aiim

Predictive modelling of survival and length of stay in critically ill patients using sequential organ failure scores

Rein Houthooft[a,*], Joeri Ruyssinck[a], Joachim van der Herten[a], Sean Stijven[a], Ivo Couckuyt[a], Bram Gadeyne[a,b], Femke Ongenae[a], Kirsten Colpaert[b], Johan Decruyenaere[b,c], Tom Dhaene[a], Filip De Turck[a]

[a] Department of Information Technology (INTEC), Ghent University – iMinds, Gaston Crommenlaan 8, B-9050 Ghent, Belgium
[b] Department of Intensive Care Medicine, Ghent University Hospital, De Pintelaan 185, 2 K12 IC, B-9000 Ghent, Belgium
[c] Department of Internal Medicine, Ghent University, De Pintelaan 185, B-9000 Ghent, Belgium

ARTICLE INFO

Article history:
Received 12 September 2014
Received in revised form 8 December 2014
Accepted 20 December 2014

Keywords:
Mortality prediction
Length of stay modeling
Support vector machines
Critical care
Sequential organ failure score

ABSTRACT

Introduction: The length of stay of critically ill patients in the intensive care unit (ICU) is an indication of patient ICU resource usage and varies considerably. Planning of postoperative ICU admissions is important as ICUs often have no nonoccupied beds available.
Problem statement: Estimation of the ICU bed availability for the next coming days is entirely based on clinical judgement by intensivists and therefore too inaccurate. For this reason, predictive models have much potential for improving planning for ICU patient admission.
Objective: Our goal is to develop and optimize models for patient survival and ICU length of stay (LOS) based on monitored ICU patient data. Furthermore, these models are compared on their use of sequential organ failure (SOFA) scores as well as underlying raw data as input features.
Methodology: Different machine learning techniques are trained, using a 14,480 patient dataset, both on SOFA scores as well as their underlying raw data values from the first five days after admission, in order to predict (i) the patient LOS, and (ii) the patient mortality. Furthermore, to help physicians in assessing the prediction credibility, a probabilistic model is tailored to the output of our best-performing model, assigning a belief to each patient status prediction. A two-by-two grid is built, using the classification outputs of the mortality and prolonged stay predictors to improve the patient LOS regression models.
Results: For predicting patient mortality and a prolonged stay, the best performing model is a support vector machine (SVM) with $G_{A,D} = 65.9\%$ (area under the curve (AUC) of 0.77) and $G_{S,L} = 73.2\%$ (AUC of 0.82). In terms of LOS regression, the best performing model is support vector regression, achieving a mean absolute error of 1.79 days and a median absolute error of 1.22 days for those patients surviving a nonprolonged stay.
Conclusion: Using a classification grid based on the predicted patient mortality and prolonged stay, allows more accurate modeling of the patient LOS. The detailed models allow to support the decisions made by physicians in an ICU setting.

1. Introduction

1.1. Problem statement

The patient length of stay (LOS) is often seen as an indication of the patient resource usage in the intensive care unit (ICU) [1]. Currently ICU physicians generally plan only a single day ahead based on clinical judgement. Automated scheduling assistance based on patient survival and LOS predictions would be beneficial in optimizing ICU resource usage, e.g., estimating the number of occupied beds, as well as individualized patient care. Moreover, this enables the adaptation of surgery scheduling to the predicted ICU load. In addition, predictive ICU models could be a building block in the larger process of making *do not resuscitate* (DNR) decisions to determine whether to stop patient therapy to avoid unnecessary suffering and treatment costs.

In this work, machine learning techniques are trained based on the sequential organ failure (SOFA) score [2–5], a score assessing

* Corresponding author. Tel.: +32 9 331 49 42.
E-mail address: rein.houthooft@intec.ugent.be (R. Houthooft).

http://dx.doi.org/10.1016/j.artmed.2014.12.009
0933-3657/

the daily individual degree of organ failure. The SOFA score is an objective score that allows for calculation of both the number and the severity of organ dysfunction in six organ systems (respiratory, coagulation, liver, cardiovascular, renal, and neurological). The score can measure individual or aggregate organ dysfunction over time and is useful to evaluate morbidity. Although the SOFA scoring was not developed to predict outcome, the obvious relationship between organ dysfunction and mortality has been demonstrated in several studies [3,6,7].

Moreover, patient mortality and LOS estimation is studied in a live monitoring setting by taking into account not only data from the first few days after admission, but also from a moving data window. This allows us to predict the status for a patient with an arbitrary current LOS. Additionally, our models assign a degree of certainty to their classification outputs, allowing ICU physicians to adapt their interpretation of the model to its credibility.

1.2. Related work

In previous studies, ICU patient mortality and LOS modelling has been conducted by taking into account patient data only from day one [8,9]. These studies generally focus on determining whether a patient will have a prolonged stay, i.e., a LOS crossing some predefined threshold. [10] apply machine learning models trained on monitored data from the first five days after patient admission, to predict the patient prolonged LOS, using a 350,000 patient dataset. Contrary to their approach we examine the use of SOFA scores as well as raw data for ICU modelling purposes. SOFA scores are used in a dynamic Bayesian network setting by Sandri et al. [11] to predict sequences of organ failures in a dataset of 79 critically ill patients, however they focus on predicting sequences of organ failures rather than the patient LOS or mortality. Meyfroidt et al. [12] have applied Gaussian processes in ICU patient LOS modelling. They focus on information monitored in the first 4 h after admission and focus on LOS prediction of 960 patients undergoing cardiac surgery. Silva et al. [13] also make use of SOFA scores to build predictive ICU models using a 4425 patient dataset, however their goal is to predict individual organ failures rather than patient mortality, prolonged stay and LOS. Furthermore, [14] have applied a variety of machine learning techniques to model ICU patient survival for a dataset of approximately 1623 patients. However, they focus on a specific patient subset which prevents straightforward generalization of their results.

1.3. Paper organization

The remainder of this paper is structured as follows. In Section 2, we elaborate on the applied predictive models and feature selection methods. Section 3 describes the data used for the applied modelling techniques and sets forth the SOFA score. Hereafter, Section 4 outlines the conducted experiments as well as their results, after which these are discussed in Section 5. Finally, in Section 6 general conclusions are highlighted.

2. Predictive modelling

The survival as well as the prolonged stay prediction are modelled by classification techniques, while the numeric patient LOS is modelled via regression. In this work the following methods are used for classification: artificial neural networks (ANNs) [15], k-nearest neighbors (k-NN) [16], support vector machines (SVMs) [17], classification trees (CART) [18], random forests (RF) [19] and adaptive boosting (AdaBoost) [20]. For regression we use: ANNs, k-NN, RF, support vector regression (SVR) [17], Relevance Vector Regression (RVR) [21] and regression trees (CART) [18]. Some of the experiments are executed using models implemented by SUMO Toolbox [22]. To select the most relevant features both backward elimination and RF, as an importance ranker, are used. In the following paragraphs these applied modelling techniques are described briefly.

2.1. Support vector machines

Support vector machines (SVMs) [17] are sparse kernel machines, a type of models that rely only on a subset of data, the support vectors, to predict unknown values. Additionally, they allow the use of kernels which allow the projection of input data to a different, possibly higher-dimensional space. The model separates the input data by means of a good-fitting hyperplane into two classes. Kernels can be used to transform this hyperplane into a nonlinear input separator, making it a very effective classifier. The SVMs used in this work have the following tunable parameters: a cost term C that controls the misclassification tolerance and acts as a regularization parameter, and one or more kernel parameters.

2.1.1. Probabilistic SVMs

On top of predicting a class, we would like our models to assign a probability, a belief, that a sample is classified correctly. This is done by means of a probabilistic extension of the SVM [23]. As such, a probability

$$P(y = 1|\mathbf{x}) \tag{1}$$

is given for each prediction. This is achieved by – next to optimizing the hyperplane decision boundary – fitting a sigmoid function

$$P(y = 1|\mathbf{x}; A, B) = \frac{1}{1 + \exp(Ay(\mathbf{x}) + B)} \tag{2}$$

on the decision values y of the SVM classifier. Herein the parameters A and B are estimated by running a maximum likelihood algorithm for Eq. (2) over the original training set.

2.2. Support vector regression

Support vector regression (SVR) [17] is the application of SVMs to regression tasks, in which a linear function is fit through the training set. In this work ϵ-SVR is used, which builds a tube around the fitted curve in which the data points have a zero cost value. Doing so allows us to fit a curve in such a way that many points reside inside this tube. Again, the predictions only depend on a subset of data, the support vectors, which lie on the tube boundaries. Also, kernels can be used to transform the linear fit to a nonlinear curve. The parameters used are the radius ϵ of the tube, which controls the tolerance towards deviation from the fitted curve and acts as a regularization parameter, and one or more kernel parameters.

2.3. Relevance vector machines

SVMs require cross-validation in order to optimally tune their parameters. Furthermore, they cannot capture output uncertainty naturally. Relevance vector machines (RVMs) [21] resemble SVMs, but apply a Bayesian approach to learning by introducing a prior distribution of the SVM weights. They are also sparse as most of the posterior weight distributions concentrate around zero and are hence negligible. The nonzero weights, called relevance vectors, are, unlike SVMs, not based on their distance to a hyperplane or tube. Furthermore, they require less parameter tuning than SVMs, but it can be computationally expensive to train them on large datasets. The regression version of the RVM is called Relevance Vector Regression (RVR).