Artificial Intelligence in Medicine 61 (2014) 45–52

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

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

Sepsis mortality prediction with the Quotient Basis Kernel

Vicent J. Ribas Ripoll [a,*], Alfredo Vellido [b], Enrique Romero [b], Juan Carlos Ruiz-Rodríguez [c]

[a] Centre de Recerca Matemàtica, Campus de Bellaterra, Edifici C, 08193 Bellaterra (Barcelona), Spain
[b] Soft Computing (SOCO) Research Group, Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya, Edifici Omega, Campus Nord, 08034 Barcelona, Spain
[c] Shock Organ Dysfunction and Resuscitation (SODIR) Research Group, Critical Care Service, Vall d'Hebron University Hospital, Vall d'Hebron Research Institute, Universitat Autònoma de Barcelona, Passeig de la Vall d'Hebron, 119-129, 08035 Barcelona, Spain

ARTICLE INFO

Article history:
Received 7 April 2013
Received in revised form 12 March 2014
Accepted 16 March 2014

Keywords:
Kernels
Support vector machines
Sepsis
Mortality prediction
Critical care

ABSTRACT

Objective: This paper presents an algorithm to assess the risk of death in patients with sepsis. Sepsis is a common clinical syndrome in the intensive care unit (ICU) that can lead to severe sepsis, a severe state of septic shock or multi-organ failure. The proposed algorithm may be implemented as part of a clinical decision support system that can be used in combination with the scores deployed in the ICU to improve the accuracy, sensitivity and specificity of mortality prediction for patients with sepsis.
Methodology: In this paper, we used the Simplified Acute Physiology Score (SAPS) for ICU patients and the Sequential Organ Failure Assessment (SOFA) to build our kernels and algorithms. In the proposed method, we embed the available data in a suitable feature space and use algorithms based on linear algebra, geometry and statistics for inference. We present a simplified version of the Fisher kernel (practical Fisher kernel for multinomial distributions), as well as a novel kernel that we named the Quotient Basis Kernel (QBK). These kernels are used as the basis for mortality prediction using soft-margin support vector machines. The two new kernels presented are compared against other generative kernels based on the Jensen–Shannon metric (centred, exponential and inverse) and other widely used kernels (linear, polynomial and Gaussian). Clinical relevance is also evaluated by comparing these results with logistic regression and the standard clinical prediction method based on the initial SAPS score.
Results: As described in this paper, we tested the new methods via cross-validation with a cohort of 400 test patients. The results obtained using our methods compare favourably with those obtained using alternative kernels (80.18% accuracy for the QBK) and the standard clinical prediction method, which are based on the basal SAPS score or logistic regression (71.32% and 71.55%, respectively). The QBK presented a sensitivity and specificity of 79.34% and 83.24%, which outperformed the other kernels analysed, logistic regression and the standard clinical prediction method based on the basal SAPS score.
Conclusion: Several scoring systems for patients with sepsis have been introduced and developed over the last 30 years. They allow for the assessment of the severity of disease and provide an estimate of in-hospital mortality. Physiology-based scoring systems are applied to critically ill patients and have a number of advantages over diagnosis-based systems. Severity score systems are often used to stratify critically ill patients for possible inclusion in clinical trials. In this paper, we present an effective algorithm that combines both scoring methodologies for the assessment of death in patients with sepsis that can be used to improve the sensitivity and specificity of the currently available methods.

1. Introduction

Sepsis is a clinical syndrome defined by the presence of both infection and systemic inflammatory response syndrome (SIRS). Sepsis can lead to severe sepsis, which is defined by organ dysfunction or an even more severe condition, septic shock and multi-organ failure [1].

This pathology has clearly increased over the last 20 years, rising to 750,000 cases per year in the United States of America alone. As the population ages and treatment becomes more aggressive, this figure is likely to grow [2,3]. In western health systems, patients with sepsis account for as high as 25% of bed utilisation in the

* Corresponding author. Tel.: +34 680435939.
E-mail address: vribas@crm.cat (V.J. Ribas Ripoll).

http://dx.doi.org/10.1016/j.artmed.2014.03.004
0933-3657/

intensive care unit (ICU), and the pathology occurs in 1–2% of all hospitalisations. The mortality rates of sepsis are very high, ranging from 12.8% for sepsis to 45.7% for septic shock [4].

These figures alone justify the need for a quantitative approach to predict mortality due to sepsis in the ICU. The extreme demands of this clinical environment further require prediction methods that are both robust and feasible within the constraints of a busy ICU.

In this paper, we describe a novel sepsis mortality prediction method that embeds the available data in a suitable feature space and uses algorithms based on linear algebra, geometry and statistics for inference. More specifically, we present a novel kernel for multinomial distributions, the Quotient Basis Kernel (QBK), which is based on the re-parametrisation of the input space through algebraic geometry and algebraic statistics. This kernel can be efficiently modelled algebraically by means of the regular exponential family. In addition, we present a generative approach that exploits the inner structure of our data to build a set of efficient, closed-form kernels that are best suited for multinomial distributions.

The QBK is the result of calculating the covariance of the design matrix of a Gröbner basis. In this paper, we hypothesise that the QBK is particularly well suited for predicting sepsis-related mortality. Not only does it exploit the inner structure of the data (i.e., it is generative), but it also provides a geometric representation accounting for the inner dependencies between its inputs [5] (in our case, these inputs are the Sequential Organ Failure Assessment (SOFA) and Simplified Acute Physiology Score (SAPS) at ICU admission).

This representation is closely related to graphical models [6] in such a way that these kernels could be considered *open-box* methods.

We compared the performance of the proposed QBK method for the prediction of mortality due to sepsis to methods obtained using a number of alternative kernels in the well-known multiparameter intelligent monitoring in intensive care (MIMIC II) database using soft-margin support vector machines (SVMs) [7]. We also compared the QBK to a standard method used in clinical practice that is based on the basal SAPS score [8] (i.e., through the automatic selection of a threshold) and logistic regression.

The paper is organised as follows: Section 2 presents the database used in this study along with the two main indices used for mortality prediction: the SOFA and SAPS scores. In Section 3, we describe a simplified version of the Fisher kernel for multinomial families. This section also provides an overview of the kernels based on the Jensen–Shannon metric [9], with a special emphasis on reparametrisation of the log-Laplace transform term of a regular exponential family. The section closes with the formal definition of the novel QBK and a short overview of SVMs. In Section 4, we show the experimental prediction results for each different kernel and their comparison with standard mortality prediction based on the basal SAPS score.

2. Background

In normal clinical practice, clinicians often treat severely ill patients in the later stages of sepsis or in its more severe manifestations. In many cases, these patients may be suffering from a combination of chronic and acute disease.

Illness scoring systems are commonplace in the ICU. The rationale for using these systems in the clinical environment is to ensure that the increased complexity of disease in patients currently being treated is consistently represented and assessed. A specific goal of severity scoring systems is to use the representative attributes of each patient to describe the relative risks they face and identify where the patient can be located along the continuum of illness severity.

It is increasingly evident that the ultimate goal of severity scoring is more than just obtaining an overall figure representing the degree of physiological disturbance. Severity scoring can be used in conjunction with other risk factors, such as disease aetiology, to anticipate and estimate outcomes such as ICU mortality. These estimates can be calculated at the time that a patient presents for care or at the time of entry into a clinical trial. Therefore, scoring systems can serve as a pre-treatment protocol. Moreover, they can also be updated during the course of therapy to describe the course of illness and provide an alternative for the evaluation of treatment response.

2.1. Sequential Organ Failure Assessment Score

In 1994, the European Society of Intensive Care Medicine (ESICM) [10] organised a consensus meeting in Paris to create the SOFA score, with the aim of objectively and quantitatively describing the degree of organ dysfunction/failure over time in groups of patients or individuals. The following represent the two main applications of this system:

1. Improving the understanding of the natural history of organ dysfunction/failure and the interrelation between the failure of various organs/systems.
2. Assessing the effect of new therapies on the course of organ dysfunction/failure to characterise patients at admission into the ICU (and even serve as an ICU entry criterion) or evaluate treatment efficacy.

Originally, the SOFA score was not designed to predict outcome (mortality) but to describe a series of complications of the critically ill. Although any assessment of morbidity is related to mortality to some extent, the SOFA score was not designed to describe organ dysfunction/failure according to mortality. However, and as described elsewhere [11], a SOFA score greater than 7 has important ICU outcome prediction capabilities. Moreover, when combined with additional parameters, it provides a very powerful set of predictors not only for outcome assessment but also for the study of the evolution of sepsis into its more severe states.

SOFA limits the number of organs/systems to six: respiratory (inspiration air pressure), coagulation (platelet count), liver (bilirubin), cardiovascular (hypotension), central nervous system (Glasgow coma score), and renal (creatinine or urine output). The scoring for each organ/system ranges from 0 for *normal function* to 4 for maximum *failure/dysfunction*. The final SOFA score is the addition of the dysfunction indexes for all organs/systems. Therefore, the maximum possible SOFA score is 24, corresponding to maximum failure for all six organs/systems.

From a clinical perspective, a SOFA score greater than 1 corresponds to multiple organ dysfunction syndrome (MODS), and cardiovascular SOFA scores greater than 2 correspond to septic shock. Normally, SOFA scores are calculated at ICU admission. However, daily calculations of SOFA scores (dynamic SOFA) [12,13] provide valuable information regarding organ dysfunction evolution and prognosis. To expedite the calculation of the Gröbner bases presented below, the input values for SOFA have been transformed into deciles before calculating all kernels in the reported experiments.

2.2. Simplified Acute Physiology Score for ICU patients

The SAPS uses 14 routinely measured biologic and clinical variables [8] to develop a simple scoring system to calculate the risk-of-death (ROD) in ICU patients. Each variable is assigned a range from 0 to 4 (i.e., the score ranges from 0 to $14 \times 4 = 56$).