Support Vector Machine Classification Applied on Weaning Trials Patients

B.F. Giraldo
Universitat Politècnica de Catalunya, Spain

A. Garde
CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Spain

C. Arizmendi
Universitat Politècnica de Catalunya, Spain

R. Jané
Institut de Bioenginyeria de Catalunya (IBEC), Spain

I. Diaz
Hospital de la Santa Creu i Sant Pau, Spain

S. Benito
Hospital de la Santa Creu i Sant Pau, Spain

INTRODUCTION

The most common reason for instituting mechanical ventilation is to decrease a patient's work of breathing. Many attempts have been made to increase the effectiveness on the evaluation of the respiratory pattern by means of respiratory signal analysis. This work suggests a method of studying the lying differences in respiratory pattern variability between patients on weaning trials. The core of the proposed method is the use of support vector machines to classify patients into two groups, taking into account 35 features of each one, previously extracted from the respiratory flow. 146 patients from mechanical ventilation were studied: Group S of 79 patients with Successful trials, and Group F of 67 patients that Failed on the attempt to maintain spontaneous breathing and had to be reconnected. Applying a feature selection procedure based on the use of the support vector machine with leave-one-out cross-validation, it was obtained 86.67% of well classified patients into the Group S and 73.34% into Group F, using only eight of the 35 features. Therefore, support vector machines can be an interesting classification method in the study of the respiratory pattern variability.

Support vector machine (SVM) is a binary classifier based on a supervised statistical learning through examples (Cristianini, 2000). SVM introduced by Vapnik (1998) and studied by others research (Burges, 1998; Veropoulos, Cristianini, & Campbell, 1999) is a new and powerful learning methodology that can deal mainly with nonlinear classification and regression. SVM are based on the use of decision hyperplanes which determine decision boundaries. A decision hyperplane is one that separates the set of objects having different class memberships. The algorithm is simple enough to be analyzed mathematically, since it can be shown that correspond to a linear method in a high-dimensional feature space, nonlinearly related to input space (Hearst, 1998).

One of the most challenging problems in intensive care is the process of discontinuing mechanical ventilation. Critical-care clinicians must carefully weight the benefits of rapid liberation for mechanical ventilation against the risks of premature trials of spontaneous breathing and extubation. The need for accurate prediction applies to all phases of weaning, beginning with reductions in mechanical support, as patients are increasingly able to support their own breathing, followed by trials of unassisted breathing, which often precede extubation, and ending with extubation (Meade, Guyatt, Cook, & Griffith, 2001). When mechanical ventilation is discontinued, up to 25 percent of patients have respiratory distress severe enough to require reinstitution of ventilation support (Tobin, 2001).

The respiratory pattern describes the mechanical function of the pulmonary system and can be characterized by the following time series: inspiratory time (T_I), expiratory time (T_E), breath duration (T_{Tot}), tidal volume (V_T), fractional inspiratory time (T_I/T_{Tot}), mean inspiratory flow (V_T/T_I), and frequency-tidal volume ratio (f/V_T).

The aim of the present work is the analysis of the respiratory pattern variability in patients during weaning trials by means of support vector machines, in order to find differences between patients that had been capable to maintain spontaneous breathing, and patients that failed in that purpose.

The following method employs techniques from the area of signal processing and incorporates the doctor's knowledge, in order to achieve a satisfactory level of

reliability so as to act as a decision support system in respiratory treatments. This research is an effort to exploit the capabilities of SVM as a classification method to identify and analyze the respiratory pattern variability.

ANALYZED DATA

Respiratory flow was measured in 146 patients on weaning trials from mechanical ventilation (WEANDB data base). These patients were recorded in the Departments of Intensive Care Medicine at Santa Creu i Sant Pau Hospital and Getafe Hospital, according to a protocol approved by the local ethic committees. Using clinical criteria based on the T-tube test, the patients were classified into two groups: Group S, 79 patients with Successful trials after 30 minutes and Group F, 67 patients that Failed in the purpose to maintain spontaneous breathing, and had to be reconnected after 30 minutes of weaning trials.

Respiratory flow was obtained using a pneumotachograph connected to an endotraqueal tube. The pneumotach consists in a Datex – Ohmeda monitor with a Validyne Model MP45-1-871 Feature-Reluctance Transducer. The signal was recorded at a sampling frequency of 250 Hz during 30 minutes.

From each recorded signal, the aforementioned time series were obtained: inspiratory time (T_I), expiratory time (T_E), breath duration (T_{Tot}), tidal volume (V_T), fractional inspiratory time (T_I/T_{Tot}), mean inspiratory flow (V_T/T_I), and frequency-tidal volume ratio (f/V_T)

METHODOLOGY

Data Preprocessing

Each one of the seven-time series was processed by moving a Running Window (RW), with a width range of 15, the best width ($p < 0.001$), consisting of several consecutive breath cycles. Five statistics were calculated for each window: the mean ($\bar{x}$), standard deviation (S), kurtosis (K), skewness (Sk), and interquartile range (IQR) of the value. In this way, 35 new time series were obtained for each patient.

Next, the data of each patient were analyzed independently by applying a k-means clustering algorithm, which automatically determines the number of clusters.

In this study, for all patients, there was a main cluster containing the most part of the patterns (i.e., data points) with a considerable internal cohesion (low intra-cluster variance), corresponding to more than 96%, for each group. The patients were distributed in 80% for training process, and 20% for test process.

Leave-One-Out Cross-Validation

Leave-one-out cross-validation is a suitable technique to give a rough estimate of the performance of learning algorithm (SVM), and can be used to value the accuracy of the classifier (Ambroise & McLachlan, 2002). This external cross-validation should be undertaken to correct the error committed by choosing randomly the training and testing set. In this cross-validation process, each sample is removed in turn and the classifier is built using the remaining samples. The class of the removed sample is then predicted using the classifier created each case. This process is repeated for each sample in the original training set with which the classification accuracy is estimated.

The 35 features were ranked through a forward selection procedure, and beginning with the most relevant feature, the cross-validation error was estimated in turn adding the next relevant feature each time.

The best result was obtained with the next eight features: $S(T_E)$, $\bar{x}(T_{Tot})$, $\bar{x}(T_I)$, $\bar{x}(T_E)$, $S(T_{Tot})$, $IQR(T_{Tot})$, $\bar{x}(f/V_T)$, $\bar{x}(T_I/T_{Tot})$. Only these most relevant features were used in the final classification process, based on support vector machines.

Histogram Equalization

In order to achieve a reduction of the overlap between successful and unsuccessful patients (Group S and Group F), the variances of the features might be similar. Unfortunately, these variances cannot always be expected to be similar, so to solve this problem, an equalization of the histograms of the previously selected features as a nonaffine normalization process is proposed in this article (Hilger & Ney, 2001).

Histogram equalization or cumulative distribution function equalization is a general nonparametric method to make the cumulative distribution function (CDF) of some given data match to a reference distribution (Balchandran & Mammone, 1998).

The principle of this method is to find a nonlinear transformation to reduce the mismatch of two signals.

Related Content

Digital Systems Innovation for Health Data Analytics
Kamaljeet Sandhu (2020). *Handbook of Research on Optimizing Healthcare Management Techniques (pp. 261-270).*
www.irma-international.org/chapter/digital-systems-innovation-for-health-data-analytics/244710

The Application of Intelligent Keywords to Patient Adherence Management System
Yin-Pin Yangand Tsai-Ya Lai (2013). *International Journal of E-Health and Medical Communications (pp. 102-119).*
www.irma-international.org/article/the-application-of-intelligent-keywords-to-patient-adherence-management-system/107057

Health Systems Simulation
David Lyell, Rosemarie Sadsadand Andrew Georgiou (2008). *Encyclopedia of Healthcare Information Systems (pp. 646-655).*
www.irma-international.org/chapter/health-systems-simulation/12996

No Silver Bullet: Identifying Security Vulnerabilities in Anonymization Protocols for Hospital Databases
Nan Zhang, Liam O'Neill, Gautam Das, Xiuzhen Chengand Heng Huang (2012). *International Journal of Healthcare Information Systems and Informatics (pp. 48-58).*
www.irma-international.org/article/silver-bullet-identifying-security-vulnerabilities/75149

Application of Handheld Computing and Mobile Phones in Diabetes Self-Care
Abdullah Wahbeh (2013). *Cases on Healthcare Information Technology for Patient Care Management (pp. 254-275).*
www.irma-international.org/chapter/application-handheld-computing-mobile-phones/73953