Medical Informatics in a United and Healthy Europe
K.-P. Adlassnig et al. (Eds.)
IOS Press, 2009

doi:10.3233/978-1-60750-044-5-590

Dynamic Data Analysis and Data Mining for Prediction of Clinical Stability

Kristien VAN LOON [a], Fabian GUIZA [b], Geert MEYFROIDT [c], Jean-Marie AERTS [a], Jan RAMON [b], Hendrik BLOCKEEL [b], Maurice BRUYNOOGHE [b], Greet VAN DEN BERGHE [c], Daniel BERCKMANS [a,1]

[a] *Division Measure, Model & Manage Bioresponses, Katholieke Universiteit Leuven, Belgium*
[b] *Department of Computer Sciences, Katholieke Universiteit Leuven, Belgium*
[c] *Department of Intensive Care Medicine, Katholieke Universiteit Leuven, Belgium*

Abstract. This work studies the impact of using dynamic information as features in a machine learning algorithm for the prediction task of classifying critically ill patients in two classes according to the time they need to reach a stable state after coronary bypass surgery: less or more than nine hours. On the basis of five physiological variables different dynamic features were extracted. These sets of features served subsequently as inputs for a Gaussian process and the prediction results were compared with the case where only admission data was used for the classification. The dynamic features, especially the cepstral coefficients (aROC: 0.749, Brier score: 0.206), resulted in higher performances when compared to static admission data (aROC: 0.547, Brier score: 0.247). In all cases, the Gaussian process classifier outperformed logistic regression.

Keywords. Gaussian processes, time series analysis, intensive care, binary probabilistic classifier, cepstral coefficients

1. Introduction

In cardiac surgery it would be very helpful to have a system that provides an early alert if there is a high probability that a patient will be disconnected from ventilation during the next day since this would lead to a more optimal planning in the ICU. It was shown that alterations in vital signals are relevant to patient management [1], so we wanted to use the trends of some of those vital signals during the first hours of ICU stay to predict a short or prolonged length of stay from early on. All living organisms are characterised by the fact that they are complex individually different time-variant and dynamic (so called CITD systems) [2]. Consequently, it is expected that taking these characteristics into account will lead to better models of the physiological signals of intensive care patients. So far, univariate and multivariate autoregressive analyses as well as the calculation of the cepstrum of physiological variables have been applied in several medical studies [3, 4] to analyze individual patients. For making classifications using many variables at the same time, several data mining techniques are available. However, in most cases no dynamic information about the patients is taken into

[1] Corresponding Author: Division Measure, Model & Manage Bioresponses, Katholieke Universiteit Leuven, Kasteelpark Arenberg 30, 3001 Leuven, Belgium; E-mail: daniel.berckmans@biw.kuleuven.be.

account when applying the data mining approach. Several attempts on temporal feature extraction for time series classification have been made [5]. The objective of this study was to investigate whether the prediction of the timeframe in which the minimal clinical conditions to start weaning of the mechanical ventilation are reached, can be more accurately predicted by using dynamic information of the individual patients when compared to predictions on the basis of static admission data.

2. Materials and Methods

2.1. Data Generation

In the surgical ICU of the university hospital of Leuven, 22 beds are reserved for cardiac surgery patients. We screened the data of all patients admitted to the ICU after planned coronary bypass surgery between February 2006 and December 2006. We selected five physiological variables that were routinely monitored in these patients (Philips Merlin monitor) and saved with a sample interval of 1 minute in the Patient Data Management System (Metavision®, iMD-Soft®) to be used as inputs: heart rate (bpm), systolic arterial blood pressure (mmHg), systolic pulmonary pressure (mmHg), blood temperature (°C) and oxygen saturation. Data of a total of 203 patients was analysed. For these patients also admission data was used (see Table 1).

2.2. Modelling Analysis

2.2.1. Abstraction of Dynamic Information

Besides the mean and standard deviations of the signals (Avgstd), we used the technique explained in this section.

Table 1. The population description

	Class 1 (Criteria met<=9h)	Class 2 (Criteria met>9h)
Number of patients	102	101
Age (mean ± std)	66 ± 11	66 ±10
Sex (male / female) %	66% / 34%	84% / 16%
BMI (mean ± std)	27.7 ± 4.6	27.7 ± 3,8
Normal lung function (%)	90%	84%
Diabetes (%)	65%	68%
Creatinine (mg/dL) (mean ± std)	1.15 ± 0.45	1.18 ± 0.40
NYHA class (I/II/III/IV) (%)	55% / 28% / 12% / 5%	61% / 19% / 19% / 1%

First, an autoregressive (AR) model was used which has the following equation for the multivariate case (MAR):

$$Y(t) = \sum_{m=1}^{M} A(m)Y(t-m) + E(t) \qquad (1)$$

Every observation is made up of a linear combination of M prior observations (the order of the model) and a white noise term. $Y(t) = \left[y_1(t), y_2(t), ..., y_K(t) \right]$ is the vector of simultaneously measured values at time t for K variables and $E(t)=\left[e_1(t),e_2(t),...,e_K(t)\right]$ is a prediction error vector [6].

Secondly, the cepstrum, defined as the inverse Fourier transform of the short-time logarithmic amplitude spectrum [7] can summarize the dynamic information of a time series. More detailed, the real cepstrum for a sequence x is given by the sequence y:

$$y=\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left|X(e^{j\omega})\right|e^{j\omega t}d\omega \tag{2}$$

Where $X(e^{j\omega})$ is the Fourier transform of y. Cepstral coefficients (CEP) decay rapidly to zero, so only the first few coefficients are needed to capture most of the dynamic information in the signal.

2.2.2. Gaussian Processes for Classification

Gaussian processes (GP), a type of kernel method, are a machine learning technique that has been successfully used to model and forecast real dynamic systems [8]. In probabilistic binary classification the task is to determine for an unlabeled test input vector the probability of belonging to a given class when a training set is given. The training set is comprised of training input vectors and their corresponding binary class labels (+1 if the input vector belongs to the class, –1 otherwise). In the remainder of this text the input vectors will be referred to as *examples*.

In GP binary classification, the GP over a function is transformed through a logistic function so that its outputs lie in the $[0,1]$ interval, and can be thus interpreted as probabilities. Since the GP is defined by its covariance function, which in turn is defined by a set of parameters, then training the GP amounts to finding the values of the parameters such that the probability of the data given these parameters is maximized. In this study these parameters are determined through the use of expectation propagation [9]. The covariance function used in this study is the Rational Quadratic with ARD (automatic relevance determination) [8].

2.3. Protocol

According to the protocol in our ICU, the following criteria have to be met before sedation can be switched off: hemodynamic stability (dobutamine ≤ 5 µg/kg/min, levophed ≤ 0,2µg/kg/min and lactate < 2 mmol/L), respiratory stability (the oxygen saturation in arterial blood flow (PaO_2) ≥ 75 mmHg, the fraction of inspired oxygen concentration (FiO_2) ≤ 0.5, the positive end-expiratory pressure (PEEP) ≤ 8 mbar), temperature stability (blood temperature > 36°C, peripheral temperature > 30°C) and blood loss stability (sum of blood loss of all drains < 100ml/h). The considered task was restated as follows: Predict the probability that the patient will begin to satisfy the stability criteria within each of the following time frames (classes): class 1: earlier than nine hours after admission; class 2: later than nine hours after admission. This nine hour threshold was chosen such that the resulting classes contained roughly the same amount of patients. This division also conforms to an intuitive classification used by intensivists into patients that recover quickly and those that require prolonged ICU stays. Data from each patient, collected during the first four hours ICU stay, were used to generate the different time-series models, the parameters of which were used as the features of the examples. One of the two possible class labels was assigned to each example. The types of examples (input vectors) used to train the GP classifiers in our experiments are *Signal Average and standard deviation*: Each example is a 20 dimensional vector containing four values for each of the five physiological variables. For two intervals of two hours the mean and the standard deviation were calculated for each signal; *MAR coefficients*: All five physiological variables were used as input of a first order MAR model. So, matrix A of Eq. (1) was a 5×5 matrix resulting in a 25 dimensional vector; and *Cepstral coefficients (CEP)*: Each example contained the five (CEP_5) first cepstral coefficients of each physiological variable. This resulted in a 25 dimensional vector.

Training examples for each classifier were labelled positive (+1) if the moment when the patient became stable started within the first nine hours after admission and were labelled negative (–1) otherwise. All examples generated for all patients from one type of time series model and their corresponding class labels were collected in one dataset and a stratified 10-fold cross-validation was performed. The examples assigned to each fold were selected randomly but the class distribution per fold was kept equal to that of the complete dataset. The obtained probabilities allowed for the computation of an aROC (area under the receiver operating characteristics curve) for each classifier. To evaluate the calibration of the predicted probabilities the Brier Score [10] was computed.

3. Results and Discussion

Table 2 gives the obtained aROCs as well as the Brier scores for each experiment with the GP. The middle column contains the results obtained when using a logistic regression (LOGREG) model, included here as a baseline for performance. The increase in performance for all GP models versus the LOGREG models was found to be significant, except for the model based on admission. So, although logistic regression techniques are commonly used in medical applications, other classifiers might lead to better results. This was, among others, also concluded by Sakai et al. and Erol et al. [11, 12].

All dynamic models perform better than the model purely based on admission information, with respect to both the Brier score and aROC. The GP with five cepstral coefficients (CEP_5) had the best performance (lowest Brier score and highest aROC). The difference in performance is shown to be significant. This is in agreement with our assumptions that it is a promising approach towards feature extraction for time-series prediction tasks. The poor performance of the models based on static information alone can be attributed to the similarity of these parameters for the two classes in our particular population (see Table 2). There is no significant statistical difference in performance between any of the logistic regression models with respect to the Brier score. When considering the aROC's, Avgstd has the best performance with statistical significance.

To improve the generalization of the classifiers it would also be of use to increase the number of patients used during training. This would allow the use of more features, possibly resulting in better performances while still avoiding over-fitting.

Table 2. Overview of the results

aROC / Brier Score	LOGREG	GP
Admission (7)	0.543/ 0.249	0.547 / 0.247
Avgstd (20)	0.628 / 0.241	0.713 / 0.214
MAR (25)	0.591 / 0.250	0.708 / 0.219
CEP_5 (25)	0.542 / 0.247	0.749 / 0.206

To our knowledge, the work of Verduijn et al. [5] is most closely related to our study. The main conclusion of their work was that induction of numerical meta-features is preferable to extraction of symbolic meta-features using existing clinical concepts, a result that complements our findings.

4. Conclusion

In this study, the use of dynamic information, obtained from physiological signals in various ways, was investigated for the prediction of stability of ICU patients, i.e., the moment that the weaning of the mechanical ventilation can be started. The main conclusion of this work is that it is preferable to use dynamic information of the first few hours after admission in the ICU over using only static admission data for the considered prediction task. All models based on dynamic information performed better with respect to aROC's and Brier scores. When compared to logistic regression, the Gaussian process classifier results in a better performance in all cases.

References

[1] Rivera-Fernandez, R., Nap, R., Vazquez-Mata, G. et al. (2007) Analysis of physiologic alterations in intensive care unit patients and their relationship with mortality. *Journal of Critical Care* 22:120–128.
[2] Quanten, S., De Valck, E., Mairesse, O. et al. (2006) Individual and time-varying model between sleep and thermoregulation. *Journal of Sleep Research* 15:243–244.
[3] Wada, T., Akaike, H., Yamada, Y. et al. (1988) Application of multivariate autoregressive modeling for analysis of immunological networks in man. *Computers & Mathematics with Applications* 15:713–722.
[4] Curcie, D.J., Craelius, W. (1997) Recognition of individual heart rate patterns with Cepstral vectors. *Biological Cybernetics* 77:103–109.
[5] Verduijn, M., Sacchi, L., Peek, N. et al. (2007) Temporal abstraction for feature extraction: A comparative case study in prediction from intensive care monitoring data. *Artificial Intelligence in Medicine* 41(1):1–12.
[6] Box, G.E., Jenkins, G.M., Reinsel, G.C. (1994) *Time Series Analysis: Forecasting and Control.* Prentice-Hall International, New Jersey.
[7] Rangayyan, R.M. (2002) *Biomedical Signal Analysis: A Case Study Approach*. Wiley Interscience, New York.
[8] Seeger, M. (2004) Gaussian processes for machine learning. *International Journal of Neural Systems* 14(2):69–106.
[9] Minka, T.P. (2001) *A Family of Algorithms for Approximate Bayesian Inference*. PhD Thesis, Masachusetts Institute of Technology, USA.
[10] Hand, D.J. (1997) *Construction and Assessment of Classification Rules*. John Wiley & Sons, Chichester.
[11] Sakai, S., Kobayashi, K., Toyabe, S.I. et al. (2007) Comparison of the levels of accuracy of an artificial neural network model and a logistic regression model for the diagnosis of acute appendicitis. *Journal of Medical Systems* 31:357–364.
[12] Erol, F.S., Uysal, H., Ergün, U. et al. (2005) Prediction of minor head injured patients using logistic regression and MLP neural network. *Journal of Medical Systems* 29(3):205–215.