Engineering Applications of Artificial Intelligence 26 (2013) 984–996

Contents lists available at SciVerse ScienceDirect

Engineering Applications of Artificial Intelligence

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

Time series classification for the prediction of dialysis in critically ill patients using echo state networks

Femke Ongenae [a,*], Stijn Van Looy [b], David Verstraeten [c], Thierry Verplancke [d], Dominique Benoit [d], Filip De Turck [a], Tom Dhaene [a], Benjamin Schrauwen [c], Johan Decruyenaere [d]

[a] Department of Information Technology (INTEC), Ghent University - Interdisciplinary Institute for Broadband Technology (IBBT), Gaston Crommenlaan 8 bus 201, B-9050 Ghent, Belgium

[b] Department of Environmental Modelling, Flemish Institute for Technological Research (VITO), Boeretang 200, 2400 Mol, Belgium

[c] Department of Electronics and Information Systems (ELIS), Ghent University, Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium

[d] Department of Intensive Care Medicine, Ghent University Hospital, De Pintelaan 185 - 2 K12 IC, B-9000 Ghent, Belgium

ARTICLE INFO

Article history:
Received 15 February 2012
Received in revised form
25 July 2012
Accepted 20 September 2012
Available online 17 October 2012

Keywords:
Time series
Classification
Echo state network
Dialysis
Feature extraction and selection

ABSTRACT

Objective: Time series often appear in medical databases, but only few machine learning methods exist that process this kind of data properly. Most modeling techniques have been designed with a static data model in mind and are not suitable for coping with the dynamic nature of time series. Recurrent neural networks (RNNs) are often used to process time series, but only a few training algorithms exist for RNNs which are complex and often yield poor results. Therefore, researchers often turn to traditional machine learning approaches, such as support vector machines (SVMs), which can easily be set up and trained and combine them with feature extraction (FE) and selection (FS) to process the high-dimensional temporal data. Recently, a new approach, called echo state networks (ESNs), has been developed to simplify the training process of RNNs. This approach allows modeling the dynamics of a system based on time series data in a straightforward way.

The objective of this study is to explore the advantages of using ESN instead of other traditional classifiers combined with FE and FS in classification problems in the intensive care unit (ICU) when the input data consists of time series. While ESNs have mostly been used to predict the future course of a time series, we use the ESN model for classification instead. Although time series often appear in medical data, little medical applications of ESNs have been studied yet.

Methods and material: ESN is used to predict the need for dialysis between the fifth and tenth day after admission in the ICU. The input time series consist of measured diuresis and creatinine values during the first 3 days after admission. Data about 830 patients was used for the study, of which 82 needed dialysis between the fifth and tenth day after admission. ESN is compared to traditional classifiers, a sophisticated and a simple one, namely support vector machines and the naive Bayes (NB) classifier. Prior to the use of the SVM and NB classifier, FE and FS is required to reduce the number of input features and thus alleviate the curse dimensionality. Extensive feature extraction was applied to capture both the overall properties of the time series and the correlation between the different measurements in the time series. The feature selection method consists of a greedy hybrid filter-wrapper method using a NB classifier, which selects in each iteration the feature that improves prediction the best and shows little multicollinearity with the already selected set. Least squares regression with noise was used to train the linear readout function of the ESN to mitigate sensitivity to noise and overfitting. Fisher labeling was used to deal with the unbalanced data set. Parameter sweeps were performed to determine the optimal parameter values for the different classifiers. The area under the curve (AUC) and maximum balanced accuracy are used as performance measures. The required execution time was also measured.

Results: The classification performance of the ESN shows significant difference at the 5% level compared to the performance of the SVM or the NB classifier combined with FE and FS. The NB+FE+FS, with an average AUC of 0.874, has the best classification performance. This classifier is followed by the ESN, which has an average AUC of 0.849. The SVM+FE+FS has the worst performance with an average AUC of 0.838. The computation time needed to pre-process the data and to train and test the classifier is significantly less for the ESN compared to the SVM and NB.

* Corresponding author. Tel.: +32 9 331 49 38; fax: +32 9 331 48 99.
E-mail addresses: Femke.Ongenae@intec.UGent.be (F. Ongenae), Stijn.VanLooy@vito.be (S. Van Looy), David.Verstraeten@UGent.be (D. Verstraeten), Thierry.Verplancke@UGent.be (T. Verplancke), Dominique.Benoit@UGent.be (D. Benoit), Filip.DeTurck@intec.UGent.be (F. De Turck), Tom.Dhaene@UGent.be (T. Dhaene), Benjamin.Schrauwen@UGent.be (B. Schrauwen), Johan.Decruyenaere@UGent.be (J. Decruyenaere).

0952-1976/$ - see front matter
http://dx.doi.org/10.1016/j.engappai.2012.09.019

Conclusion: It can be concluded that the use of ESN has an added value in predicting the need for dialysis through the analysis of time series data. The ESN requires significantly less processing time, needs no domain knowledge, is easy to implement, and can be configured using rules of thumb.

1. Introduction

Time series are a special kind of input data to machine learning problems. Most modeling techniques have been designed with a static data model in mind and are not suitable for coping with the dynamic nature of time series. Most dynamic data models are very complex in both design and training algorithms. Examples of such models based on artificial neural networks are the hidden control neural network (Levin, 1993), the neural prediction model (Iso and Watanabe, 1991), the linked predictive neural network (Tebelskis et al., 1990) and the adaptive time-delay neural network (Xie et al., 2006). Recurrent neural networks (RNNs) are often used (Robinson, 1994) since this type of artificial neural network can represent high-dimensional non-linear temporal data. Hidden Markov models (Rabiner, 1989) and neural network-hidden Markov model hybrids (Graves and Schmidhuber, 2005; Trentin and Gori, 2003) are also used to model time series data. An obstacle when using RNNs is that only a few training algorithms exist which are complex and often yield poor results (Haykin, 1994; Jaeger, 2002b).

More recently, three approaches to simplify the training process of RNNs were independently developed. These approaches are liquid state machines (LSM) (Maass et al., 2002), echo state networks (ESN) (Jaeger, 2001), and backpropagation decorrelation (BPDC) (Steil, 2006). The underlying idea of these three methods is similar and nowadays they are referred to as *reservoir computing* (Verstraeten et al., 2007). Reservoir computing has become a vivid research field and recently a special issue of "*neural networks*" was dedicated to it (Jaeger et al., 2007).

The key idea in reservoir computing is that the dynamic system producing the time series data is modeled in a *reservoir* consisting of a RNN. The reservoir is then read by a linear readout function, which is illustrated in Fig. 1. The output of this readout function can then be used to make several kinds of predictions. The training algorithm only affects the linear readout function. For training linear functions many algorithms exist such as linear regression (Fisher, 1925).

The goal of this study is to verify whether the use of reservoir computing methods is an added value in classification problems in the intensive care unit (ICU) when the input data consists of time series. We select a case study that is easily characterized by medical experts. This medical classification problem is then handled using reservoir computing, which can directly cope with time series data, and the performance is compared to more traditional machine learning approaches, which cannot directly cope with this high-dimensional temporal data and thus need to be combined with feature extraction (FE) and selection (FS) to process the time series.

LSMs and ESNs are the two pioneering reservoir computing methods. However, the two methods have a very different background (Jaeger et al., 2007). The initial ESN publications were framed in settings of machine learning and non-linear signal processing applications (Jaeger, 2001, 2002a,b, 2003; Jaeger and Haas, 2004). In contrast, LSMs were developed from a computational neuroscience background, aiming at elucidating principal computational properties of microcircuits (Maass et al., 2002, 2003, 2004; Natschläger et al., 2002).

This difference in background also explains the main difference between LSMs and ESNs (Lukoševičius and Jaeger, 2009). ESNs standardly use simple sigmoid neurons or leaky integrator neuron models, while LSMs use more sophisticated and biologically realistic models built from a spiking neuron model called the leaky integrate and fire (LIF) neuron (Maass and Bishop, 2001) and dynamic synaptic connection models (Gerstner and Kistler, 2002) in the reservoir. Since the model of both the connections and the neurons themselves in LSMs is quite sophisticated, it has a large number of free parameters to be set, which is done manually, guided by biologically observed parameter ranges. The parameters of ESNs, e.g., the warm-up drop and the leak rate, are more intuitive and can easily be set by using rules of thumb or performing parameter sweeps. Moreover, LSMs require pulse trains as input data. Translating continuous data, of which the training data of the medical problem under study in this research

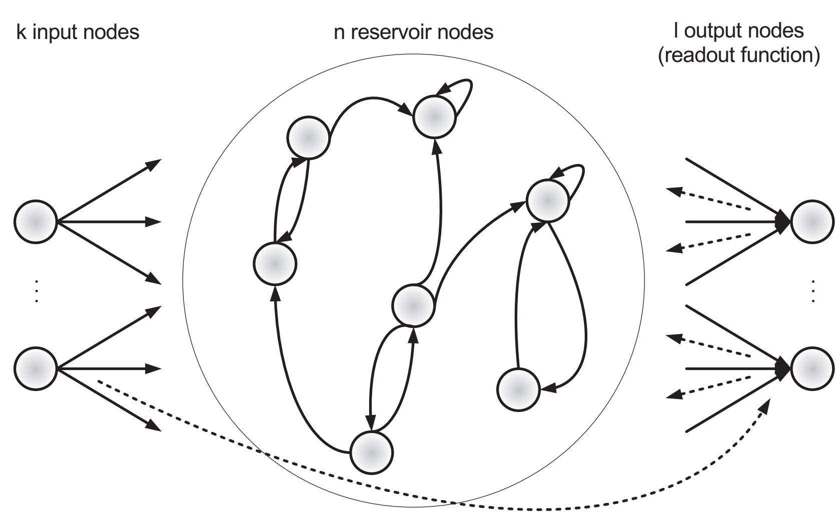

Fig. 1. The general layout of an echo state network. Circles represent input, reservoir, and output nodes. Arrows represent non-zero weighted connections. Dotted arrows denote optional connections.