Artificial Intelligence in Medicine xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

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

Using echo state networks for classification: A case study in Parkinson's disease diagnosis

Stuart E. Lacy[a,*], Stephen L. Smith[a], Michael A. Lones[b]

[a] University of York, York, United Kingdom
[b] Heriot-Watt University, Edinburgh, United Kingdom

ARTICLE INFO

Article history:
Received 12 July 2017
Received in revised form 1 February 2018
Accepted 8 February 2018

Keywords:
Parkinson's disease
Echo state networks
Neurodegenerative disease

ABSTRACT

Despite having notable advantages over established machine learning methods for time series analysis, reservoir computing methods, such as echo state networks (ESNs), have yet to be widely used for practical data mining applications. In this paper, we address this deficit with a case study that demonstrates how ESNs can be trained to predict disease labels when stimulated with movement data. Since there has been relatively little prior research into using ESNs for classification, we also consider a number of different approaches for realising input–output mappings. Our results show that ESNs can carry out effective classification and are competitive with existing approaches that have significantly longer training times, in addition to performing similarly with models employing conventional feature extraction strategies that require expert domain knowledge. This suggests that ESNs may prove beneficial in situations where predictive models must be trained rapidly and without the benefit of domain knowledge, for example on high-dimensional data produced by wearable medical technologies. This application area is emphasized with a case study of Parkinson's disease patients who have been recorded by wearable sensors while performing basic movement tasks.

1. Introduction

Reservoir computing is a general approach for modelling complex dynamical systems using a large recurrent neural network (RNN), with only the network output weights being trained [1]. Echo state networks (ESNs) are a well known implementation where the output connections are fitted using simple ordinary least-squares regression [2]. Owing to this, ESNs are significantly faster and more scalable than many existing more complex machine learning approaches and are ideally suited for time series analysis. Nevertheless, despite this significant advantage, they have yet to be widely used in data mining applications.

Many medical applications have a need for predictive models that can capture the complexity of biological disease pathways to facilitate personalised healthcare. A good example is the Parkinson's disease (PD) case study considered in this paper. PD is a debilitating progressive neurodegenerative disease that presents with a broad spectrum of movement disorders, which even expert clinicians can find challenging to characterise and discriminate from other related diseases [3]. Wearable sensors can provide significant benefits to patient care by objectively measuring movement disorders in high resolution and therefore help monitor disease progress and their use is becoming increasingly widespread [4].

ESNs, with their ability to model dynamical processes, would seem like a sensible candidate for modelling such data and provide two primary benefits. The first is that they can directly model the raw time series to identify any patterns in the underlying dynamics of the signal that conventional feature extraction techniques may miss. Their second advantage is their rapid training speed, resulting from having a closed form solution. This is an important consideration for applied predictive modelling, owing to the need to train and evaluate candidate models on a range of data sets when performing model selection and evaluation.

In this paper, we consider how ESNs can be applied to the problem of diagnosing PD from movement data of the kind that might be collected using wearable accelerometers. Since there has been little existing work in this area, we focus on exploring the key issue of how inputs and outputs can be mapped to the ESN methodology, and how this affects the predictive accuracy of the model. One aspect that is investigated is whether to segment the data before inputting into the model. This has ramifications for subsequent work on analysis of data recorded from wearable sensors by facilitating simpler processing and analysis at the cost of adding

* Corresponding author.
E-mail address: stuart.lacy@york.ac.uk (S.E. Lacy).

https://doi.org/10.1016/j.artmed.2018.02.002
0933-3657/

Please cite this article in press as: Lacy SE, et al. Using echo state networks for classification: A case study in Parkinson's disease diagnosis. Artif Intell Med (2018), https://doi.org/10.1016/j.artmed.2018.02.002

more design time [5]. To evaluate the practicality of the resulting network, a two-fold comparison is performed. First, ESN classifiers are compared to models built on summary features derived using the guidance of an expert in movement disorders, in order to establish whether ESNs offer a more flexible alternative without compromising on accuracy. The second comparison is against previous attempts on this data set, highlighting the ability of ESNs to rapidly fit an accurate model comparable with those produced from complex optimisation routines requiring significantly longer computational time.

The rest of the paper is organised as follows: Section 2 provides details of ESNs and previous applications to both classification tasks and medical problems in general, and Section 3 details the data collection process of the Parkinson's disease movement data. The experimental methodology is laid out in Section 4, while Section 5 presents the results. Finally, Section 6 concludes.

2. Echo state networks

2.1. Background

In recent years, a new and increasingly well researched dynamical systems approach to modelling complex time series has been developed, termed *reservoir computing*. As the name implies, the model is focused around what is known as a *reservoir*: a coupled system of non-linear functional elements in which dynamical behaviour can be modelled. Data is passed directly into the reservoir through a set of input nodes, while the output at each time step is determined by a linear readout. The functional elements are typically computational models of neurons as used by artificial neural networks (ANNs); for this reason reservoir computing is considered a sub-field of ANN research. An additional defining characteristic of reservoir computing approaches is that only the reservoir readout mechanism is trained—typically using ordinary least squares—allowing for a much simpler and less computationally intensive training pipeline than found in other dynamical system modelling approaches. Two common implementations of reservoir computing are Echo State Networks (ESNs) [2] and liquid state machines (LSMs) [6]. ESNs typically employ a sparsely connected reservoir of sigmoidal nodes, while LSMs use a more biologically plausible neuron model by incorporating spiking neurons in the reservoir [7]. In this paper, only ESNs are considered owing to their more efficient construction and training method.

2.2. Network configuration

ESNs comprise three distinct sets of neurons: inputs, the reservoir itself, and the output readout nodes, as shown in Fig. 1. When constructing a network, the three weight matrices $\boldsymbol{W^{in}}$, $\boldsymbol{W}$, and

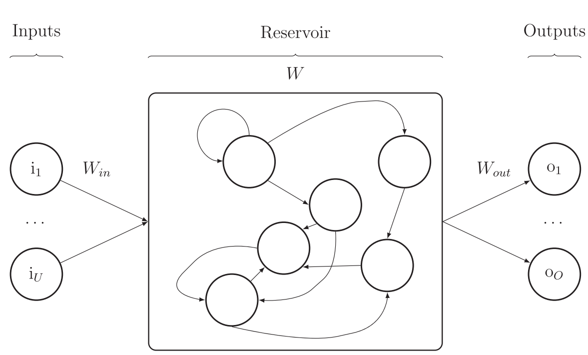

Fig. 1. The reservoir of an ESN is sparsely connected with recurrent connections providing a means of maintaining state. Only the weight matrix $\boldsymbol{W^{out}}$ is optimised during training. The number of nodes in the reservoir is much larger than shown.

$\boldsymbol{W^{out}}$ are initialised randomly, with weights typically drawn uniformly from $[-1, 1]$. The reservoir weights are further scaled by a parameter called the spectral radius, in order to fulfil what is known as the *echo state property* [8], whereby traces of previous inputs are visible as *echoes* in each following reservoir state with a diminishing presence. The reservoir is sparsely connected to allow for a co-existing range of diverse dynamics, with only around 1% of all connections non-zero. Execution of the network is governed by two state-space equations. Eq. (1) defines how the reservoir states $x(n)$ are updated at each time step, where $u(n)$ represents network inputs and $f(.)$ is a predefined activation function, typically the logistic function. By allowing a term corresponding to previous activation states, the network maintains a memory of past inputs, and so can also be used as a generative model.

$$x(n) = f(\boldsymbol{W^{in}}u(n) + \boldsymbol{W}x(n-1)) \quad (1)$$

$$y(n) = g(\boldsymbol{W^{out}}[x(n); u(n)]) \quad (2)$$

The reservoir output at a given time step $y(n)$ is governed by Eq. (2), where the concatenated vector $[x(n);u(n)]$ is often referred to as the *extended state vector* $e(n)$, and a second activation function $g(.)$ is employed. Stacking the vectors across all the time-points produces matrices $\boldsymbol{Y}$ and $\boldsymbol{E}$, as in Eq. (3). Estimates for $\boldsymbol{W^{out}}$ can be determined by solving ordinary least squares regression of the target signal $\boldsymbol{D}$ as a function of $\boldsymbol{E}$. Since there is a closed form solution for ordinary least squares, the training time is extremely fast compared to algorithms that use iterative techniques such as gradient descent.

$$Y = g(\boldsymbol{W^{out}E}) \quad (3)$$

ESNs offer several advantages over traditional recurrent neural networks (RNNs), mostly related to their simpler design and implementation. For example, most of the design choices when developing an ESN are concerned with the reservoir itself, including its size and the spectral radius parameter. Rather than having to consider multiple layers of functional elements, there is just a single set of active nodes, with the input and outputs providing interfaces to and from the reservoir. The second main difference with more established recurrent networks such as the Elman or Jordan models is that only the weights of the output nodes are modified during training, thereby providing a much simpler training procedure than those used for standard RNN techniques such as backpropagation through time.

However, ESNs have not seen as much uptake as could be expected given their simplicity and promising results on benchmark data sets. This may be a result of concerns about their theoretical understanding; the reservoir itself is commonly viewed as a black box, with little understanding on how the dynamics present in the input signal are being modelled. In addition, there is little guidance available on how to best tune the network parameters for a particular application, with optimal design often arising as a result of trial-and-error [9,10]. Ozturk et al. [11] cite this as a major limitation of ESN applicability, along with the argument that the primary hyper-parameter in ESN design—spectral radius—is not well correlated with network goodness-of-fit. Nevertheless, while ESNs have not been largely used in machine learning applications, they have found some value in neuroscience as models of the brain [12].

2.3. Applications

Despite the limitations associated with ESN construction described above, the field has demonstrated strong modelling capabilities when applied to problems with a temporal element. In particular they have demonstrated a strong ability to forecast chaotic time series, with considerable success on the well-known Mackey-Glass benchmark equation [13,9]. In addition, they have