ARTICLE IN PRESS

Neurocomputing xx (xxxx) xxxx–xxxx

Contents lists available at ScienceDirect

Neurocomputing

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

Optimizing the echo state network based on mutual information for modeling fed-batch bioprocesses

Heshan Wang[a], Chunjuan Ni[b], Xuefeng Yan[b,*]

[a] College of Electrical Engineering, Zhengzhou University, Zhengzhou 450001, China

[b] Key Laboratory of Advanced Control and Optimization for Chemical Processes of Ministry of Education, East China University of Science and Technology, Shanghai 200237, PR China

A R T I C L E I N F O

Communicated by Yang Tang

Keywords:
Echo State Network
Mutual Information
Pruning Algorithm
Penicillin cultivation process modeling
Parameters optimizing

A B S T R A C T

Echo state networks (ESNs) have become one of the most effective dynamic neural networks because of its excellent fitting performance in real-valued time series modeling tasks and simple training processes. The original ESN concept used randomly fixed created reservoirs, and this concept is considered to be one of its main advantages. However, ESNs have been criticized for its randomly created connectivity and weight parameters. Determining the appropriate weight parameters for a given task is an important problem. An optimization method based on mutual information (MI) is proposed in this study to optimize the input scaling parameters and the structure of ESN to address the aforementioned problem and improve the performance of ESN. The MI optimization method mainly consists of two parts: First, the scaling parameters of multiple inputs are adjusted based on the MI between the network inputs and outputs. Second, some output weight connections are pruned for optimization based on the MI between reservoir states. Furthermore, three MI-ESN models are proposed for a fed-batch penicillin fermentation process. Our experimental outcomes reveal that the obtained MI-ESN models outperform the ESN models without optimization and other traditional neural networks.

1. Introduction

A feature of batch processes is the disposal of crude materials for a limited period to transform basic materials into outcomes. Batch bioprocess operations have been used to generate products that have high value additions in biological, chemical, pharmaceutical, and many other growth processes. As such, monitoring the batch bioprocess operation is crucial. The features of batch biological processes involve the non-steady-state condition, strong nonlinearity, strong time-varying condition, batch-to-batch altering, and unpredictability caused by floating crude materials [1], As such, an ideal model is important for the complex modeling and control of fed-batch biological processes. Effective process predictive models for online process monitoring play primary roles in supplying truthful and reliable online measurement and analysis with regard to the quality of key products or environmental change factors because they ease prevention, and therefore, mitigate the riskiness of processes [2]. Moreover, such predictive models can significantly improve product quality.

Soft sensors are software-based sophisticated monitoring systems that have gained credit and credibility in the academic and industrial environments. Liu et al. developed a functioning nonlinear soft sensor system that can be used to assist online modeling of batch processes based on a just-in-time application method [3]. By contrast, Pierantonio et al. argued in favor of a methodology for the automated correction of misbehaviors of partial least squares (PLS) predictive modeling in fed-batch processing [4]. Based on a multistage operational method, Yu [5] proposed a robust multi-way Gaussian mixture model (MGMM) that can be used to arrange online soft sensors of the fed-batch penicillin fermentation process. The computed results proved that the proposed MGMM method significantly outperforms the extensively accredited kernel PLS approach. Then, Yu [6] proposed a support vector regression (SVR) system combined with Bayesian inference (BI) theory and showed that the BI-SVR model can outperform the single SVR model when measurement predictions are needed.

Recurrent neural networks (RNNs) have recently been successfully employed in addressing complexities in the measurement of temporal, nonlinear, time-varying, and uncertainty-related operations [7–10]. Given the dynamic mechanism and ability of RNNs to process temporally related information [11], such methods have been extensively employed for complicated easily changing operations, such as fed-batch processes.

Echo state networks (ESNs) are RNNs with a vast untrained

* Correspondence to: P.O. BOX 293, MeiLong Road NO. 130, Shanghai 200237, PR China.
E-mail address: xfyan@ecust.edu.cn (X. Yan).

http://dx.doi.org/10.1016/j.neucom.2016.11.007
Received 7 January 2016; Received in revised form 1 November 2016; Accepted 6 November 2016
Available online xxxx
0925-2312/

Please cite this article as: Wang, H., Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.11.007

recurrent part that features scattered connections and simple linear readouts. ESNs are one of the most well-known types of RNNs because of their excellent performance when nonlinear dynamic system modeling and chaos time series predictions are conducted. In ESNs, an extensive and elaborate dynamic reservoir is built to collect several features of inputs flowing through the supplying source, namely, the reservoir, toward the output readout map. Prior to the establishment of the ESN, the internal neuron topology connectivity and input scaling parameter weights are generated randomly beforehand. A reservoir may be stated as valid whenever it meets the requirement of a necessary circumstance of the state dynamics referred to as the echo state quality: The internal neuron state is an "echo" of the entire input data stream from its commencement. The ESNs have generated successful results when applied to different sequential domains, such as time series predictions [12,13], fed-batch bioprocesses modeling [7], wind speed forecasting [14] and speech processing [15]. Many variations of the original ESNs can be found in literature; for example, hybrid circle reservoir ESN [16], leaky integrator ESN [17], and filter neurons with delay & sum readout [18].

Reservoir properties (topology or connections) and weight parameters are important for learning performance. A proper understanding of reservoir dynamics and measures for these dynamics is important [19,20]. However, ESN has several limitations, such as few reservoir properties, which can hardly be comprehended; a number of manual parameters, which need tailored adjustments; and brute-force searching, which aims to maximize the efficiency of the ESN models, for instance, the size and spectral radius of internal neurons and the scaling parameters of input data. The most prominent problem of the ESN is its high-complexity random parameters. The influence of hierarchical and structured topologies is not yet sufficiently understood. Nonetheless, reservoirs require numerous trials, even luck, as several parameters, such as the connectivity of internal neuron topology and network parameter weights, are generated stochastically beforehand. The stochastic topology connectivity of internal units and network parameter weights will hardly be precisely sufficient to establish a model and will equally hardly provide a clear analysis of internal unit dynamics [21]. Thus, determining the appropriate weights parameters and optimal reservoir for a given task are important.

Throughout the past periods of research, many scholars have concentrated on new methodologies to optimize the internal neuron dynamics by focusing on links between the parameters and performance of internal neurons [22–24] and by reducing the number of neurons and connectivity size [25,26]. Dutoit et al. [25] proposed a pruning method called least angle regression and an output weight regularization method called ridge regression (RR) to calculate the readout matrix and increase the generalization ability of ESN. Venayagamoorthy and Shishir [23] investigated the effects of ESN performance obtained through the variation of two specific parameters, namely, the spectral radius of the internal neuron matrix and the settling time. Nonetheless, an approach on which parameter choice ought to be based has not yet been outlined.

Mutual information (MI) has been demonstrated to be effective in neuron network to measure the relationship of neuron information. For instance, Janusz and Liang [27] proposed a method to reduce the interconnections and large number of synaptic weights in self-organized neural network by using local interconnection based on MI. Zhang et al. [28] presented an adaptive merging and splitting algorithm based on the theory of MI to design feedforward neural networks. He et al. [29] utilized a partial MI method to indentify the subset of artificial neural network (ANN) inputs. Chen and Yan [30] proposed a minimal redundancy maximal relevance-partial MI clustering approach to remove redundant hidden layer nodes of multilayer feedforward network. Besides MI method was also applied to optimize ANN by some other researchers [31–33]. By the overview above, MI can be utilized to measure the interrelation of neurons and prune the redundant information or neurons in ANN. However, to our knowledge, MI techniques for ESNs were rarely studied.

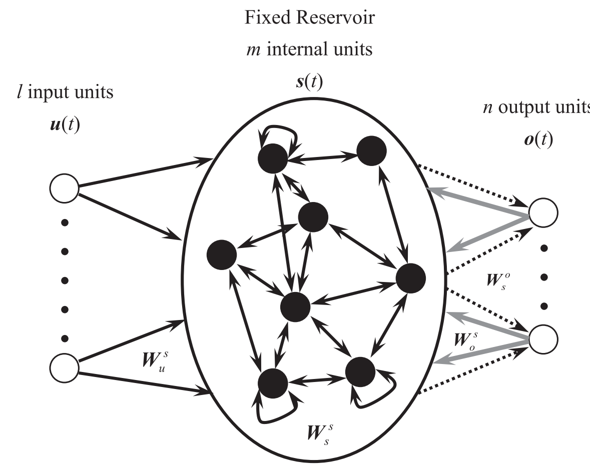

Fig. 1. The topology structure of a typical ESN. Connections that are trained in the ESN are indicated by the dashed line arrows. Feedback weight connections that are possible yet not required are indicated by the shaded line arrows. Connections that are randomly generated and adjusted in the training process are indicated by the black solid lines.

Because of the drawbacks of ESNs we described above. It is important to determine the appropriate weight parameters and optimal reservoir for a given task. This study presents a MI optimization method to optimize the input scaling and the output weight connection structure of ESN in fed-batch penicillin fermentation process. The experimental results indicate that the proposed approach exhibits superior performance in the fed-batch penicillin fermentation process. The remainder of this study is organized as follows. Section 2 presents a concise review of ESN training and design. The MI approach utilized to design and prune ESN are outlined in Section 3. A fed-batch penicillin fermentation process is discussed in Section 4. The results are discussed in Section 5. Finally, Section 6 provides a brief conclusion.

2. Echo state network and mutual information

2.1. Topology structure and training method of ESN

A typical ESN, which contains a recurrent reservoir, has l input variables, m reservoir neurons, and n output variables, as shown in Fig. 1. The state of internal neurons $\mathbf{s}(t)$ and the state of output variables $\mathbf{o}(t)$ at a specific time point t are expressed as follows [34]

$$\mathbf{s}(t) = f(\mathbf{W}_u^s \cdot \mathbf{u}(t) + \mathbf{W}_s^s \cdot \mathbf{s}(t-1) + \mathbf{W}_o^s \cdot \mathbf{o}^T(t-1)) \tag{1}$$

$$\mathbf{o}(t) = (\mathbf{W}_s^o \cdot \mathbf{s}(t))^T \tag{2}$$

where f is the internal neuron stimulation function, $\mathbf{u}(t)=[u_1(t), u_2(t), \ldots, u_l(t)]^{\mathrm{T}}$ represents the input variable, $\mathbf{s}(t)=[s_1(t), s_2(t), \ldots, s_m(t)]^{\mathrm{T}}$ is the internal neuron state, $\mathbf{o}(t)=[o_1(t), o_2(t), \ldots, o_n(t)]$ represents the output variable at a specified time step t, $\mathbf{W}_u^s$ represents the input weight matrices, $\mathbf{W}_s^s$ denotes the internal neuron connection matrices, $\mathbf{W}_o^s$ is the feedback weight matrices, $\mathbf{W}_s^o$ denotes the output weight (readout) matrices, $\mathbf{W}_u^s$, $\mathbf{W}_s^s$, $\mathbf{W}_o^s$, and $\mathbf{W}_s^o$ are $m \times l$, $m \times m$, $m \times n$, and $n \times m$ weight matrices, respectively. The initialization of reservoir state $\mathbf{s}(t)$ is a zero vector. The superscripted T represents transpose. The ESN structure without feedback connections is an open-loop topology architecture that can provide a one-step-ahead prediction for the time series problem. By contrast, the ESN structure with feedback connections is a closed-loop topology architecture that can provide an iterative prediction, which are multiple steps ahead.

$\mathbf{W}_s^s$, $\mathbf{W}_o^s$, and $\mathbf{W}_u^s$ which are generated by using the stochastic numerical values obtained from a uniform distribution, are immovable before the training process begins. In the training process of RC, only the output