ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6.
Available from http://www.i6doc.com/en/.

Quantifying the Reservoir Quality using Dimensionality Reduction Techniques

Tomas Burianek[1] and Sebastián Basterrech[2]

1- Department of Computer Science
Faculty of Electrical Engineering and Computer Science
VŠB-Technical University of Ostrava, Ostrava-Poruba, Czech Republic

2- Department of Computer Science, Faculty of Electrical Engineering
Czech Technical University, Prague, Czech Republic

Abstract. Echo State Network is a particular type of Recurrent Neural Networks that combines principles from kernels, linear regression and dynamical systems. The neural network has a random initialized hidden-hidden weights (reservoir) that keeps fixed during the training. The reservoir projects the input patterns onto a feature map. Here, we present a correlation analysis between the input space and the feature map. We use a dimensionality reduction technique (Sammon Mapping) for representing the input space. We show a correlation between the Sammon energy and the model accuracy, which can be useful for defining good reservoir topologies.

1 Introduction

During recent years the Echo State Network (ESN) has gained popularity in the Neural Network (NN) community [1]. From the point of view of the architectural design, an ESN has at least two sequential structures. First one, it is a Recurrent Neural Network (RNN) named *reservoir*. Second one, it is a memory-free supervised learning tool named *readout*. The reservoir structure has a recurrent topology composed by a large number of sigmoid neurons. The reservoir projects the input space onto a larger space. Its main role is to improve the linear separability of the original data, and to use the recurrent topology for memorizing input sequences. An identifying characteristic of the ESN model is that the hidden-hidden weights are fixed during the training process. On other words, only the parameters of the readout structure are trained. Most often, this structure is a linear regression model [1]. Several extensions of the standard ESN model have been introduced during the last years. Since 2007, these models are popularly known under the framework of Reservoir Computing (RC) [2]. In spite of the success for solving temporal learning problems, the ESN model and the RC techniques still have some well-identified limitations. Due to the reservoir weights are deemed fixed during the training process, then the reservoir initialization can impact on the final model's performance. Some attempts for initializing the reservoir have been introduced [1, 3, 4]. However, as far as we know none of them have became popularly accepted in the community.

In this short article, we present an analysis about the correlation among a representation of the input space, the feature map (projected points by the

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

reservoir) and the model accuracy. We evaluate the following work hypothesis: a "good" reservoir projects the points of the input space in a map which preserves some aspects of the topology of the input space. Hence, we evaluate the performance of a reservoir matrix that minimizes a mapping between a linear representation of the input space and the feature map. For evaluating the similarity between the input space and the feature map we use a dimensionality reduction technique named Sammon mapping [5]. We solved the dimensionality reduction problem using the Particle Swarm Optimization (PSO) metaheuristic [6, 7]. We show that exists a correlation between the Sammon energy and the ESN accuracy prediction, which can be helpful for the community in order of defining the initial reservoirs of the ESN model.

The rest of the article is organized as follows. Next section presents the ESN model and the dimensionality reduction technique. Section 3 introduces our contribution. We present the experimental results in section4. The paper ends presenting a discussion and future works.

2 Background

2.1 Formalization of the Echo State Network

Let be $\mathbf{W}^{\text{in}}$ a matrix that collects the input weights. The weights among neurons in the reservoir are collected in the reservoir matrix $\mathbf{W}$. The readout parameters are collected in a matrix $\mathbf{W}^{\text{out}}$. The reservoir state is defined as:

$$\mathbf{x}(t) = f(\mathbf{W}^{\text{in}}\mathbf{u}(t) + \mathbf{W}\mathbf{x}(t-1)), \tag{1}$$

and the ESN output is:

$$\mathbf{o}(t) = g(\mathbf{W}^{\text{out}}\mathbf{x}(t)), \tag{2}$$

where $f(\cdot)$ is a contractive function (Lipschitz function) and $g(\cdot)$ is the identity function. Only the parameters $\mathbf{W}^{\text{out}}$ presented in (2) are adjusted using a batch ridge regression [1]. The model accuracy depends of several factors that include the *Echo State Property (ESP)* [8, 9, 10], reservoir size [11], input scaling factor [12], sparsity of reservoir weights matrix [1, 11]. In the last years, several variations of the original RC models have been developed. Some models present variations in the type of topology, for example [13, 14]. Other models focus in the non-linear transformation of the input space, for instance deep layered reservoirs were analyzed in [15, 16], and other RC models have different type of reservoir neurons, for example [17, 18].

2.2 Dimensionality Reduction using the Sammon Mapping

A dimensionality reduction technique represents a set of high-dimensional points on a lower dimensional space. Each high-dimensional point is represented by a latent low-dimensional point, in such a way that the layout of the latent points is a "good" representation of the layout of the original space. Given a map $\phi(\cdot)$ between two Euclidean spaces $\mathcal{I}$ and $\mathcal{R}$, let $L_{i,j}$ be the distance between two any points i and j in $\mathcal{I}$, and let $D_{i,j}$ be the distance between the projected points in $\mathcal{R}$ (distance between $\phi(i)$ and $\phi(j)$). Sammon's mapping is a nonlinear method for dimensionality reduction of a high-dimensional data [5]. The mapped space created by the Sammon mapping preserves the main topological characteristics of the original space. We define the Sammon projection as the mapping that minimizes the following energy:

$$E = \frac{1}{\sum_{i<j} L_{i,j}} \sum_{i<j} \frac{(L_{i,j} - D_{i,j})^2}{L_{i,j}}. \tag{3}$$

3 Methodology

We consider a linear transformation of a time windows $\Delta_l(t)$ of the input sequence as $\mathbf{z}^l(t) = [\mathbf{z}(t); \mathbf{z}(t-1); \cdots; \mathbf{z}(t-l)]$, where each $\mathbf{z}(t)$ is a linear transformation of the input pattern at time t ($\mathbf{z}(t) = \mathbf{W}^{\text{in}}\mathbf{u}(t)$), and the operation $[\cdot;\cdot]$ denotes the concatenation of column vectors. The value of l is arbitrary. It can be also defined empirically studying the periodicity of the time series. We denote the input space as $\mathcal{I}$. Let $\mathcal{R}$ be the space defined by the layout of points projected by the reservoir given by expression (1). In addition, we consider the space $\mathcal{Z}$ containing the points $\mathbf{z}^l(t)$. Note that, the projected points in $\mathcal{R}$ are non-linear projections of the input sequence, and the projected points in $\mathcal{Z}$ are linear projections of the input sequence. The proposed method finds a reservoir matrix $\mathbf{W}$ such that minimizes the Sammon's energy computed between the points in $\mathcal{Z}$ and $\mathcal{R}$. As a consequence, an optimization problem emerges, the goal is to find the matrix $\mathbf{W}$ such that the Sammon energy presented in expression (3) is minimized. For our experiments we used Particle Swarm Optimization (PSO) [7]. The PSO algorithm is a popular metaheuristic for solving numerical optimization problems. The PSO algorithm is iterative, and it starts randomly initializing a set of feasible solutions. In the literature related to Swarm Intelligence, the solutions are named *particles*, and the set of points is named *swarm*. The algorithm updates the particles following some update rules [7]. In our problem, each particle represents the reservoir weights. Roughly speaking, the particles approximate to the minimum learning of the position of the best particle in the swarm, as well as they learn of their own previous steps. More details about how to use the PSO can be read in [7], and a PSO applied for improving an ESN was also studied in [19]. However, any other optimisation technique can be used for minimizing the Sammon energy. Note that, we are interested in finding a similarity between the space $\mathcal{R}$ and $\mathcal{Z}$. In this step, we do not perform the computation of the readout weights.

4 Experimental Results

We evaluate the proposal on three well-known benchmark problems. The first time series data was generated by the Lorenz attractor equations. The second benchmark dataset is the Mackey-Glass nonlinear time series (already studied in several RC articles for instance [8, 20, 21]). The last benchmark dataset is the Rossler time-series. More details about these benchmark problems can be seen in [9]. We set up the parameter of the PSO algorithm following the suggestions presented in [22, 23]. We used a swarm with 24 particles, local and global acceleration of 1.496 and inertia of 0.729. The initial ESN weights are randomly initialized using a uniform distribution in the range of $\{-0.5, 0.5\}$. We evaluate the standard ESN model with different topologies, for different combinations of reservoir size and spectral radius values $(N_r, \rho(\mathbf{W}))$. We create a grid of values $(N_r, \rho(\mathbf{W}))$ in the following domains: N_r in $\{30, 50, 100\}$ and $\rho(\mathbf{W})$ in $\{0.2, 0.5, 0.9\}$. The ESN model obtains the best accuracy when the combination $(N_r, \rho(\mathbf{W}))$ is equal to $(100, 0.9)$. We present three figures, each one contains two curves. In the top is presented a curve with the Sammon stress function computed with expression (3) according to the number of PSO iterations. In the bottom, the curve shows the evolution of mean square error (MSE) according to the PSO iterations. Figure 1 presents the results for the three benchmark problems. These figures show a correlation between the evolution of the MSEs and Sammon energy values. In addition, we present in table 1 the results of the Spearman correlation test between the evolution of MSE and Sammon stress values during the epochs of the PSO algorithm. The Spearman correlation test measures how two variables are related to each other. The null hypothesis of the test specifies that does not exist a correlation between the two data sequences. According to the Spearman rank presented in table 1 (a perfect correlation is presented when the rank is ± 1), in all the cases we reject the null hypothesis. Then, we accept the alternative hypothesis (correlation between MSE and Sammon error).

Table 1: Spearman's correlation test between the MSE and Sammon's stress error during the pre-training of SAM-ESN.

Dataset	Correlation		
	Spearman rank	Null hypothesis	Correlation type
Lorenz	0.6390832	Rejected	Strong positive
Mackey-Glass	−0.4517169	Rejected	Weak negative
Rossler	0.5884516	Rejected	Strong positive

5 Conclusions

So far as we know, it does not exists a metric for assessing the quality of the reservoirs. In general, the quality of the model is given by the accuracy of the ESN predictions (including the readout structure). In this paper, we present an attempt for assessing the reservoir quality. We consider a dimensionality reduction technique named Sammon Mapping. The quality of the reservoir is measured using the Sammon energy, which is computed using points from a linear transformation of the input space and the projected points by the reservoir.

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence
and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6.
Available from http://www.i6doc.com/en/.

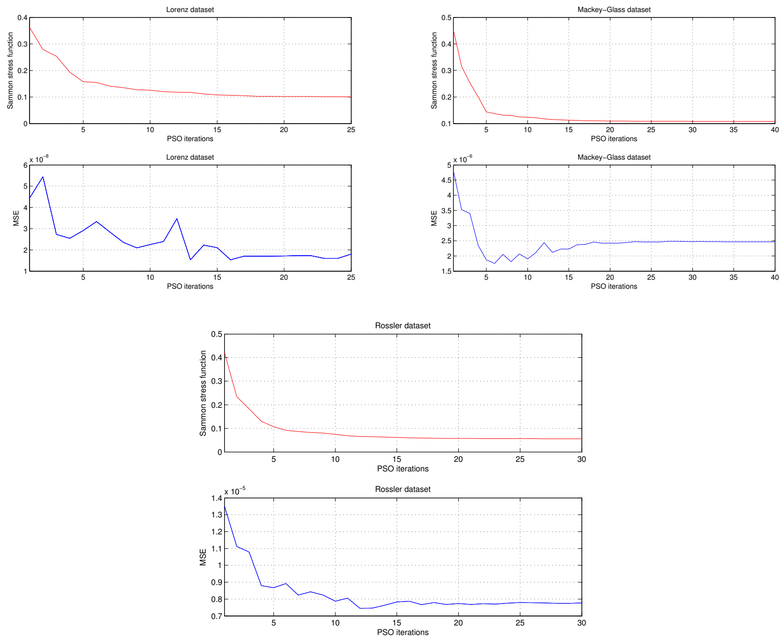

Fig. 1: Correlation between Sammon's stress function and MSE during the PSO optimization. The figures correspond to Lorenz dataset (top-left side), Mackey-Glass dataset (top-right side) and Rossler dataset (bottom).

On other words, we show with empirical results that the "good" reservoirs are the ones that project the input space on a new space with a topological similarity. We apply statistical tools for showing a correlation between the accuracy of the model and the Sammon energy, when the Sammon projection error decreases, then the MSE also decreases. In future works, we plan to analyze the performance of other dimensionality reduction techniques. Besides, we expect to analyze the correlation between Sammon energy and the RC accuracy when the reservoir has other activation functions.

References

[1] Mantas Lukoševičius and Hebert Jaeger. Reservoir Computing Approaches to Recurrent Neural Network Training. *Computer Science Review*, 3:127–149, 2009.

[2] D. Verstraeten, B. Schrauwen, M. D'Haene, and D. Stroobandt. An experimental unification of reservoir computing methods. *Neural Networks*, 20(3):287–289, 2007.

[3] Sebastián Basterrech, Colin Fyfe, and Gerardo Rubino. Self-organizing Maps and Scale-invariant Maps in Echo State Networks. In *11th International Conference on Intelligent Systems Design and Applications, ISDA 2011, Córdoba, Spain, November 22-24, 2011*, pages 94–99, November 2011.

[4] Mantas Lukoševičius. *Reservoir Computing and Self-Organized Neural Hierarchies*. PhD thesis, School of Engineering and Science. Jacobs University, december 2011.

[5] John Sammon. A Nonlinear Mapping for Data Structure Analysis. *IEEE Transactions on Computers*, C-18(5):401–409, May 1969.

[6] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of the IEEE International Conference on Neural Networks*, volume 4, pages 1942–1948, Nov 1995.

[7] M. Clerc and J. Kennedy. The Particle Swarm - Explosion, Stability, and Convergence in a Multidimensional Complex Space. *IEEE Transactions on Evolutionary Computation*, 6(1):58–73, February 2002.

[8] Herbert Jaeger. The "echo state" approach to analysing and training recurrent neural networks. Technical Report 148, German National Research Center for Information Technology, 2001.

[9] Sebastián Basterrech. Empirical analysis of the necessary and sufficient conditions of the echo state property. In *2017 International Joint Conference on Neural Networks, IJCNN 2017, Anchorage, AK, USA, May 14-19, 2017*, pages 888–896, 2017.

[10] G. Manjunath and H. Jaeger. Echo State Property Linked to an Input: Exploring a Fundamental Characteristic of Recurrent Neural Networks. *Neural Computation*, 25(3):671–696, March 2013.

[11] Mantas Lukoševičius. A Practical Guide to Applying Echo State Networks. In Grégoire Montavon, GenevièveB. Orr, and Klaus-Robert Müller, editors, *Neural Networks: Tricks of the Trade*, volume 7700 of *Lecture Notes in Computer Science*, pages 659–686. Springer Berlin Heidelberg, 2012.

[12] J. B. Butcher, D. Verstraeten, B. Schrauwen, C. R. Day, and P. W. Haycock. Reservoir Computing and Extreme Learning Machines for Non-linear Time-series Data Analysis. *Neural Networks*, 38:76–89, feb 2013.

[13] Ali Rodan and Peter Tiňo. Minimum Complexity Echo State Network. *IEEE Transactions on Neural Networks*, 22:131–144, 2011.

[14] Peter Tiño Ali Rodan. Simple Deterministically Constructed Cycle Reservoirs with Regular Jumps. *Neural Computation*, 24(7):1822–52, 2012.

[15] Claudio Gallicchio, Alessio Micheli, and Luca Pedrelli. Deep reservoir computing: A critical experimental analysis. *Neurocomputing*, 268(Supplement C):87 – 99, 2017. Advances in artificial neural networks, machine learning and computational intelligence.

[16] Claudio Gallicchio and Alessio Micheli. Echo state property of deep reservoir computing networks. *Cognitive Computation*, 9(3):337–350, Jun 2017.

[17] Wolfgang Maass. Liquid State Machines: Motivation, Theory, and Applications. In *In Computability in Context: Computation and Logic in the Real World*, pages 275–296. Imperial College Press, 2010.

[18] Sebastián Basterrech and Gerardo Rubino. Echo State Queugeing Networks: a Combination of Reservoir Computing and Random Neural Networks. *Probability in the Engineering and Informational Sciences*, pages 1–20, 2017.

[19] Sebastián Basterrech, Enrique Alba, and Václav Snášel. An Experimental Analysis of the Echo State Network Initialization Using the Particle Swarm Optimization. In *Nature and Biologically Inspired Computing (NaBIC), 2014 Sixth World Congress on*, pages 214–219, July 2014.

[20] Claudio Gallicchio and Alessio Micheli. Architectural and Markovian factors of echo state networks. *Neural Networks*, 24(5):440 – 456, 2011.

[21] Herbert Jaeger and Harald Haas. Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication. *Science*, 304(5667):78–80, 2004.

[22] M. Clerc. The swarm and the queen: towards a deterministic and adaptive particle swarm optimization. In *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406)*, volume 3, pages 19–57, 1999.

[23] R. C. Eberhart and Y. Shi. Comparing inertia weights and constriction factors in particle swarm optimization. In *Proceedings of the 2000 Congress on Evolutionary Computation. CEC00 (Cat. No.00TH8512)*, volume 1, pages 84–88, 2000.