A survey of the past, present and future of echo state networks

Guan-Fang wu[†] and Hong-Yan Cui

School of Information and Telecommunications, Beijing University of Post and Telecommunications State Key Lab. Of Networking and Switching Technology Key Lab. of network system architecture and convergence Beijing 100876, China e-mail:18713510225@163.com

Along with the development of Machine Learning, statistic and Artificial Intelligence, people are exposed to myriad of big data. Meanwhile, accurate data analysis is difficult. Echo state network (ESN) algorithms are widely researched and applied in many fields. Owing to their potential for exact prediction and simple training process, scientists pay more attention to the research of ESN. In this paper, the representative research is carried out to sum up the research achievements on ESN, and the future development direction is discussed by pointing out the key technical challenges and we suggest several strategies for tackling the challenges.

Keywords: Echo State Network; Prediction; Structure Improvement; Modelling capability analysis

1. Introduction

As the Internet becomes our most important communications infrastructure, service providers try to further improve their functionality, including security, reliability, privacy, and multiple service functionalities. The future Internet will embody a large number of objects that, through standard communication protocols and unique addressing schemes, provide information and services to the final users [1]. A large amount of data is generated in this process, and the variety and complexity of the data makes it difficult to handle. Humans are exposed to myriad of sensory data all the time and can capture their important aspects in a way that allows them to be used in the future [2].

Prediction is employed to interpret and estimate the future development trend of things, that is, reveal the inherent trends or laws of development of things and point out the future development of the approaches and results according to the past and present development of things. Prediction uses



scientific methods of analysis and comprehensive information of all aspects. An observed sequence often conveys a meaning, whereby it is difficult to decipher independent fragments of this sequence in isolation. Meaning is often inferred from events or observations that are received in adjacent time [3]. Therefore, modeling the temporal component of the observations plays a critical role in effective information representation. The prediction model is different for different time series. The initial time series forecasting method is a simple extrapolation of global fit in the known time domain; it is suitable for linear systems, and the forecast accuracy is poor.

For decades, it has been one of the key challenges in artificial intelligence research to imitate the effectiveness and robustness of the human brain. Recent neuroscience findings provide insights into the design of information systems by simulating the representation of information in animal brains. This discovery motivated the emergence of the artificial neural network models composing of many nodes and connections. Through the above ideas, scientists have proposed several artificial neural network models by imitating the structure of the human brain. The topology of such systems is either regular or random. However, the real neural networks are actually neither entirely regular nor totally random. Neural networks can be viewed as a general representation of dynamic systems [4]. Its internal structure can be seen as a black box especially for highly nonlinear system. Feedforward neural networks such as the Multilayer Perceptrons (MLPs) [5], Radial Basis Functions (RBFs) have got great success in the nonlinear system identification. Among the possible model, Recurrent Neural Network (RNN) is widely used in direct or indirect controller design. RNN is better than the feedforward neural network in terms of its dynamic memory and time embedding capabilities in dynamic system modeling. However, it is difficult to realize due to the complexity of the training in a short time. We observe that the time series will need to operate in an extremely complex context. It is unlikely to guarantee the robustness of the algorithm in the simple network environment.

Despite these constraints, in the era of big data, prediction technology is critical for scientific research, which compels researchers to continually seek ways to optimize prediction algorithm or otherwise improve their performance. For these reasons, Echo State Networks (ESNS) have been greatly developed due to their simple topology and superior prediction performance.

In this article, we carry out a summary on the present situation of ESN. Section II briefly introduces the traditional ESN structure and subsequently follows by a detailed overview of some representative improved ESN architectures in Section III. Section IV contains a brief note about the impact on application and development of ESN. The conclusion in Section V provides a



perspective of the potential impact of ESN as well as key questions that remain to be answered.

2. Echo State Networks

2.1. Introduction of ESN

Echo State Network (ESN) is a novel design of RNNs, providing an architecture and supervised learning principle, which is first proposed by Jaeger and Haas [6] from German Jacobs University. It is a type of three-layered recurrent network with a sparse and random hidden layer that is not trained. The networks possess the echo state property, that, if given a long enough sequence, the network will always eventually reach the same state, regardless of the initial state. In other words, its internal state echoes the input sequence [7]. The main idea of ESN is to drive a random, large, fixed dynamic pool with the input sequence. The internal weights of the reservoir are not changed during training process; only the reservoir-to-output connections are trained. Meanwhile, to ensure the echo state characteristics, the dynamic spectral radius is less than 1. The general structure of ESN is shown in Fig. 1.



Figure 1. The skeleton frame of ESN.

2.2. Structure and operation mechanism of ESN

In this paper, a discrete-time neural network with M input units, N internal network units, and L output units (Fig. 2) is used. The input units, internal units and output units at time step n are denoted as

$$u(n) = (u_1(n), u_2(n) \dots u_M(n))$$
(1)

$$x(n) = (x_1(n), x_2(n) \dots x_N(n))$$
(2)



$$y(n) = (y_1(n), y_2(n)...y_L(n))$$
(3)

respectively. The state equations of ESN are as follows:

$$x(n) = f(W_{in}u(n) + Wx(n-1) + W_{back}y(n-1))$$
(4)

$$y(n) = f^{out}(W_{out}(u(n), x(n), y(n-1)))$$
(5)

where $f = f(f_1, f_2, ..., f_N)$ is the internal units output functions. W_{in} and W_{back} represent input weight matrix and output feedback weight matrix, respectively. W is state matrix of reservoir and W_{out} represents the matrix of the dynamic reservoir to the output layer. $f^{out} = f(f_1^{out}, f_2^{out}, ..., f_L^{out})$ is the output function.



Figure 2. The basic structure of ESN.

During the training progress of the network, the weight matrix of the reservoir network is randomly generated and it will not change, only the output weight matrix W_{out} is required to be trained [8] so that the network can achieve convergence. Here, we compute the normalized root minimized mean square error (NRMSE) as follows to represent the training error

$$NRMSE = \sqrt{\frac{\|d(n) - y(n)\|^2}{\|y(n) - \overline{y(n)}\|^2}}$$
(6)

where d(n) is the true value of the output sequence, y(n) is the training output sequence, $\| \bullet \|$ denotes the Euclidean norm, and \bullet means the empirical



mean.

The performance of ESN has a close relationship with the dynamic features. Moreover, these dynamic features depend on the values of the ESNs' global parameters, such as the spectral radius, the number of input units, the internal excitation functions, the reservoir size and sparsity. By the definition of echo state property, we know that an ESN state is a function of finite history of inputs presented to the network-the state is the echo of the input history. In practice, when the random and sparse internal weight matrix W of ESN is scaled that its spectral radius $\rho(W)$ is smaller than 1, the echo state property is obtained. In the past few years, although this is used widely in reservoir definition, this condition is neither necessary nor sufficient to ensure the echo state property. In addition, the short-term memory (STM) [9] is an effective way to understand the properties of the ESNs.

The construction of ESN illustrated as Fig. 3 is simple, as well as the training algorithm, so the training process can be completed quickly. However, the large scale of internal neurons in ESN lead to an enormous number of training samples of the entire network, making the training progress difficult. The prediction stage is full of uncertain factors.



Figure 3. Flow chart of constructing an ESN.

3. Recently Proposed ESN Architectuees and Performance Analysis

3.1. Analysis and improvement of ESN

Aimed at the problems mentioned in Section II. B., most of the solutions have been preliminary designed.

Because the performance of ESN has a close relationship with the dynamic features which depend on the values of the ESNs global parameters. A number



of improved ESN structure algorithms have been developed over the years.

3.1.1. Analysis of the reservoir

Up to now, how to use as little as possible neurons to constitute a reserve pool with as few as possible connections, so that it can generate a more complex and diverse state space has been a focus of ESN research.

The representative work mainly includes creating the reservoir in a way similar to the "growth" of the biological neural structure. One example of the idea is in [10]. The main line of this paper is to use simple network topology and simplify levels to establish a simple mathematical model, which bears excellent practical application ability. Hence, the name is called minimum complexity echo state network.

A further step ahead of this idea is the combination of ESN and complex networks. Using the small world and the non-scale dynamic pool extension, the experimental results show that this ESN has better forecasting precision. The best solution is to define a robust dynamic pool. A similar approach is proposed in literature [11], which aims at employing lateral inhibition mechanism (LIM) in a DESN dynamic pool, obtaining mixed DESN (DMESN) that enhances the prediction accuracy, and at the same time also improves the robustness of the algorithm.

Besides, because sigmoid function is difficult to achieve in hardware, random bit sequence neurons can be used instead to build the network. Luko'sevi'cius M. et al. [12] injected leaky integrator neurons into reservoir, putting forward an alternative solution. Another interesting example is that researchers improved the structure of echo state network with wavelet functions injected, which improves the prediction performance of the ESN. Very similar to this approach is the one in literature [13], in which Wang Se realized the transformation of the original ESN to a sigmoid hybrid wavelet ESN and enlarged the ESN memory space. At the same time, they added some adjusting wavelet neurons to retain desirable nonlinear characteristics. Experimental results show that SWHESN has a more powerful development cycle, a more stable operation, and less computation. SWHESN predicts 46% higher than the ESN without typical bias, while running time is just 30% of ESN.

An interesting example of attempts to go well beyond the current vision of the state-of-the-art is introduced in [14].

Fan and Han proposed a new ESN structure, which consists of a circulating pool with fixed feedback connections. They presented an on-line designed ESN model based on the particle swarm algorithm in their paper, which can be applied to the recognition of identified systems. The particle swarm optimization (PSO) exports the weight of the ESN by the online training, not by calculation. This greatly improves the accuracy of the modeling and avoids the reciprocal operation, so its application is greatly expanded.

Scardapane et al. [15] put forward an effective algorithm which removes internal redundancy of the reservoir during the training process. The algorithm is based on the correlations of node states. Therefore, it depends only on the input signal. Rachez and Hagiwara proposed a new point of view by extending the traditional ESN architecture in [16], where examples of this new attitude can be found. The number of units required to achieve similar performance in semantic modeling is much larger than that of traditional RNNs. They propose a new architecture to extend the traditional ESN with a pre-regression feature layer and nonlinear output. By using the gradient descent with less computational complexity and automatic capturing the similarity between the statements, the structure can learn the features in a supervised way. They modify the dynamics of the network in a way that allowed it to be significantly better than a single ESN. The study of nonlinear output makes the whole network similar to a feedforward network with memory layer.

As detailed in literature [17], deterministic structured dynamic pool structure with jump period has good generalization performance. The dynamic pool weight learning of the structure is the linear output weight of the hybrid optimization strategy, and the dynamic pool weight training consists of nonlinear optimization techniques [18].

3.1.2. Stability and generalization ability

In reality, some modeling sequences have chaotic characteristics. Rohitash C. et al. [19] employ cooperative coevolution to train neural networks for solving chaotic time series problems. The idea of cooperative coevolution is to decompose a problem into subcomponents and employ evolutionary algorithms for solving them.

In addition, Limin Z. et al. [20] introduce L-curve method into ESN to eliminate the effect of ill-pose. They use Ridge regression(R-R), which contain a parameter that can be determined by the Bayesian posterior probability maximization method or the Bootstrap resampling method, instead of the original linear regression.

A delay and sum structure was introduced in [21], where a trainable delay is introduced on the connection of the output neurons. In addition, literature [22] use a linear associative memory readout called Minimum Average Correlation Energy (MACE) filter as a novel readout of ESNs for detecting neural action potentials [22].

Under the comprehensive consideration of stability and generalization as well as the type of neuron function to simplify the model and improve



performance at the same time, H. Y. Cui et al [23] proposed a novel architecture called WMCESN with wavelet neurons injected into the reservoir to replace the original sigmoid neurons on the basis of a Simple-Ring Echo State Network [24]. By using different wavelet functions on different datasets, it is demonstrated that the WMCESN model largely improves the prediction performance only with little time cost radius range and has more stable prediction ability. Moreover, the model enriches the types of memory function in dynamic pool, and improves the memory ability of dynamic pool.

3.2. Preliminary Summary

The evolution of prediction algorithm in terms of neural network technology has created a heterogeneous landscape. Table I summarizes some mainstream algorithms used in prediction.

Researcher	Time	Data Set	Method
A.S. Lapedes	1987	Logistic model	MLP
A.S. Weigend	1990	Sunspot	RBF
T. Jerome	1994	Electronic load	RNN
H. Jaeger [7]	2004	Mackey-Glass	ESN
Y. Gao	2005	Mackey-Glass and gas furnace data	FNN
S. Aydogan	2007	Mackey-Glass	MFLNN
Z.W. Shi and M. Han	2007	Chaos and the Yellow River runoff	SVESM
G. Daniel	2008	IBM stock index, sunspot, oil price	FRNN
A. Nouri	2009	Robot trajectory	HBRM
Y. Peng	2011	Mobile communication traffic	FESN
A. Rachez	2012	words in the vocabulary	ESN
L.M. Zhang	2013	Blast furnace gas quantity	LC-ESN
M. Han	2014	Lorenz System and gas furnace data	ESN-SCKF
M. Han	2015	H'enon Map and Lorenz System	FSDESN

TABLE I. NEURAL NETWORK USED FOR PREDICTION

4. ESN Applications

In view of the fact that ESN is simple modeling and has good performance. There have been several examples of ESN in a variety of application domains. In addition to the time series prediction [6, 25], there are applications in image edge detection [26], nonlinear control [27] and so on.

ATLANTIS

5. The Road Ahead

At present, artificial intelligence and deep learning based on neural network are still an important part of future development. Therefore, it is very important to improve and mature neural network. However, there remains a great deal of work to be done in improving the learning process, where current focus is on optimization of reservoir. Because the core question is that dynamic reservoir, acting like a black-box with hundreds of neurons inter-connected or selfconnected, is the most important part of ESNs. In addition, further inspiration and techniques may be found from the optimization of the readout network and the adjustment of neurons. In the scientific area there have been, and still are, challenges about ESN design and implementation.

In summary, the biggest hotspot is concentrated on further theoretical exploration and optimization of network. To study more related theory to design a more reliable and compact network so as to effectively solve practical problems, there is still a lot of work remains to be done.

6. Conclusions

Research on ESN has produced creative and novel approaches, but there are still challenges exist. In this paper, we provided a critical but constructive analysis of the field of ESN, including its representative models, which illustrates the achievements in this field as well as obstacles to its future progress. The main challenge is to improve robustness, efficiency, and accuracy of the enabling prediction algorithms of ESN. We outlined both research and application directions that could improve the capabilities and effectiveness of prediction systems.

As a field of development in full swing, how to improve algorithm theory, especially the in-depth research of connotation behind the reservoir, and how the advantages of ESN are fully reflected to explore more suitable application of ESN in practice, all of these, will be a direction that is worth research attention in the field of neural network.

Acknowledgment

This work was supported by the project supported by National Natural Science Foundation of China (61201153), the National 973 Program of China under Grant (2012CB315805), and Prospective Research Project on Future Networks in Jiangsu Future Networks Innovation Institute (BY2013095-2-16).



References

- L. Atzori, A. Iera, and G. Morabito, "From "Smart Objects" to "Social Objects": The Next Evolutionary Step of the Internet of Things," IEEE Commun. Mag., vol. 52, no. 1, pp. 97-105, Jan. 2014.
- [2] I. Arel, D. C. Rose and T. P. Karnowski, "Deep Machine Learning-A New Frontier in Artificial Intelligence Research," IEEE Computational Intelligence Magazine, Nov. 2010.
- [3] G. Wallis and H. Blthoff, "Learning to recognize objects," Trends Cogn. Sci., vol. 3, no. 1, pp. 23-31, 1999.
- [4] M. D. Skowronski, J. G. Harris, "Noise-robust automatic speech recognition using a predictive echo state network," IEEE Transactions on Audio Speech and Language Processing, vol. 15, no. 5, pp. 1724-1730, 2007.
- [5] D. Erdogmus, O. Fontenla-Romero, J. C. Principe, A. Alonso-Betanzos, and E. Castello, "Linear-least-squares initialization of multilayer perceptrons through backpropagation of the desired response," IEEE Transactions on Neural Networks, vol. 16, no. 2, pp. 325-337, 2005.
- [6] H. Jaeger, H. Haas, "Harnessing Nonlinearity: Predicting chaotic systems and saving energy in wireless communication," Science, vol. 304, no. 5667, pp.78-80, 2004.
- [7] G. K. Venayagamoorthy, "Online design of an echo state network based wide area monitor for a multimachine power system," Neural Networks, vol. 20, no. 3, pp. 404-413, 2007.
- [8] J. C. Fan, M. Han, "Online Designed of Echo State Network Based on Particle Swarm Optimization for System Identification," The 2nd Internatioal Conference on Intelligent Control and Information Processing, pp. 559-563, July 25-28, 2011.
- [9] H. Jaeger, "Short term memory in echo state networks," German Nat. Res. Center Inform. Technol., St. Augustin, Germany, Tech. Rep. GMD 152, 2002.
- [10] S. Wang, X. J. Yang and C. J. Wei, "Harnessing Non-linearity by Sigmoidwavelet Hybrid Echo State Networks (SWHESN)," IEEE Proceedings of the 6th World Congress on Intelligent Control and Automation, June 21-23, pp. 3014-3018,2006.
- [11] H. Y. Cui, X. Liu and T. J. Zhou, "Performance evaluation of new echo state networks based on complex network," Science, vol. 19, no. 1, pp. 87-93, Feb. 2012.
- [12] M. Luko'sevi'cius, D. Popovici, H. Jaeger, et al., "Time Warping Invariant Echo State Networks," IUB Technical Report No.2 International University Bremen, 2006.



- [13] M. L. Xu and M. Han, "Online Prediction for Multivariate Time Series by Echo State Networks based on Square-root Cubature Kalman Filter," Proceedings of the 33rd Chinese Control Conference July 28-30, pp. 5065-5070, 2014.
- [14] J. C. Fan and M. Han, "Online Designed of Echo State Network Based on Particle Swarm Optimization for System Identification," IEEE The 2nd International Conference on Intelligent Control and Information Processing, July 25-28, pp. 559-563, 2011.
- [15] S. Scardapane, G. Nocco, D. Comminiello, "An Effective Criterion for Pruning Reservoirs Conference in Echo State Networks," IEEE 2014 International Joint Conference on Neural Networks (IJCNN) July 6-11, 2014, pp. 1205-1212.
- [16] A. Rachez and M. Hagiwara, "Augmented Echo State Networks with a Feature Layer and a Nonlinear Readout," WCCI 2012 IEEE World Congress on Computational Intelligence, pp. 10-15, June, 2012.
- [17] F. Z. Tang, P. Tino and H. H. Chen, "Learning the Deterministically Constructed Echo State Networks," Neural Network (IJCNN), 2014 International Joint Conference, pp. 77-83, July 2014.
- [18] M. H. Yusoff and Y. C. Jin, "Modeling Neural Plasticity in Echo State Networks for Time Series Prediction," Computational Intelligence (UKCI), 2014 14th UK Workshop, pp. 1-7, Sept. 2014.
- [19] R. Chandra, M. J. Zhang, "Cooperative coevolution of Elman recurrent neural networks for chaotic time series prediciton," Neurocomputing, vol. 86, pp. 116-123, 2012.
- [20] L. M. Zhang, X. P. Guan and H. J. Yang, "Improved Ill-Posed Echo State Network and Its Application to Blast Furnace Gas Amount Forecast," Proceedings of the 32nd Chinese Control Conference, pp. 4641-4645, July 26-28,2013.
- [21] G. Holzmann and H. Hauser, "Echo State Networks with filter neurons and a delay and sum readout," Neural Netw, vol. 23, no. 2, pp. 244-256, Mar. 2010.
- [22] N. J. Dedual, M. C. Ozturk and J. C. Sanchez, "An associative memory readout in ESN for neural action potential detection[C]," Neural Networks (IJCNN) International Joint Conference, pp. 2295-2299, 2007.
- [23] H. Y. Cui and C. Feng, "Analysis of Prediction Performance in Wavelet Minimum Complexity Echo State Networks," The Journal of China University of Posts and Telecommunications, May 2012.
- [24] A. Rodan and P. Tino, "Minimum complexity echo state network," IEEE Trans. Neural Networks, vol. 22, pp. 131-144, 2011.



- [25] M. Han, Z. W. Shi, W. Guo, "Reservoir neural state reconstruction and chaotic time series prediction," Acta Phys Sin, vol. 56, no. 1, 2007: 43.
- [26] C. D. Pei, "Echo state networks and its application on image edge detection," Comput Eng Appl, vol. 44, no. 19, 2008:172.
- [27] D. M. Xu, J. Lan and J. C. Principe, "Direct adaptive control: an echo state network and genetic algorithm approach," Proceedings of the IEEE International Joint Conference on Neural Networks. Montreal, 20015: 1483.