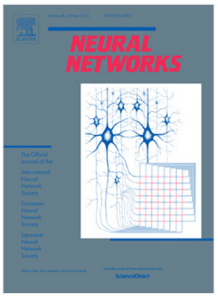
Accepted Manuscript



Echo state networks are universal

Lyudmila Grigoryeva, Juan-Pablo Ortega

PII: S0893-6080(18)30251-X
DOI: https://doi.org/10.1016/j.neunet.2018.08.025
Reference: NN 4029

To appear in: *Neural Networks*

Received date : 6 June 2018
Revised date : 26 August 2018
Accepted date : 31 August 2018

Please cite this article as: Grigoryeva, L., Ortega, J.-P., Echo state networks are universal. *Neural Networks* (2018), https://doi.org/10.1016/j.neunet.2018.08.025

Echo state networks are universal

Lyudmila Grigoryeva[1] and Juan-Pablo Ortega[2,3]

Abstract

This paper shows that echo state networks are universal uniform approximants in the context of discrete-time fading memory filters with uniformly bounded inputs defined on negative infinite times. This result guarantees that any fading memory input/output system in discrete time can be realized as a simple finite-dimensional neural network-type state-space model with a static linear readout map. This approximation is valid for infinite time intervals. The proof of this statement is based on fundamental results, also presented in this work, about the topological nature of the fading memory property and about reservoir computing systems generated by continuous reservoir maps.

Key Words: reservoir computing, universality, echo state networks, ESN, state-affine systems, SAS, machine learning, fading memory property, echo state property, linear training, uniform system approximation.

1 Introduction

Many recently introduced machine learning techniques in the context of dynamical problems have much in common with system identification procedures developed in the last decades for applications in signal treatment, circuit theory and, in general, systems theory. In these problems, system knowledge is only available in the form of input-output observations and the task consists in finding or *learning* a model that approximates it for mainly forecasting or classification purposes. An important goal in that context is finding families of transformations that are both computationally feasible and versatile enough to reproduce a rich number of patterns just by modifying a limited number of procedural parameters.

The versatility or flexibility of a given machine learning paradigm is usually established by proving its ***universality***. We say that a family of transformations is universal when its elements can approximate as accurately as one wants all the elements of a sufficiently rich class containing, for example, all continuous or even all measurable transformations. In the language of learning theory, this is equivalent to the possibility of making approximation errors arbitrarily small [Cuck 02, Smal 03, Cuck 07]. In more mathematical terms, the universality of a family amounts to its density in a rich class of the type mentioned above. Well-known universality results are, for example, the uniform approximation properties of feedforward neural networks established in [Cybe 89, Horn 89, Horn 91] in the context of static continuous and, more generally, measurable real functions.

A first solution to this problem in the dynamic context was pioneered in the works of Fréchet [Frec 10] and Volterra [Volt 30] one century ago when they proved that finite Volterra series can be

[1]Department of Mathematics and Statistics. Universität Konstanz. Box 146. D-78457 Konstanz. Germany. Lyudmila.Grigoryeva@uni-konstanz.de

[2]Universität Sankt Gallen. Faculty of Mathematics and Statistics. Bodanstrasse 6. CH-9000 Sankt Gallen. Switzerland. Juan-Pablo.Ortega@unisg.ch

[3]Centre National de la Recherche Scientifique (CNRS). France.