Stability and Topology in Reservoir Computing

Larry Manevitz and Hananel Hazan

Department of Computer Science, University of Haifa, Mount Carmel, Haifa 31905, Israel
[manevitz,hhazan01]@cs.haifa.ac.il

Abstract. Recently Jaeger and others have put forth the paradigm of "reservoir computing" as a way of computing with highly recurrent neural networks. This reservoir is a collection of neurons randomly connected with each other of fixed weights. Amongst other things, it has been shown to be effective in temporal pattern recognition; and has been held as a model appropriate to explain how certain aspects of the brain work. (Particularly in its guise as “liquid state machine”, due to Maass et al.) In this work we show that although it is known that this model does have generalizability properties and thus is robust to errors in input, it is NOT resistant to errors in the model itself. Thus small malfunctions or distortions make previous training ineffective. Thus this model as currently presented cannot be thought of as appropriate as a biological model; and it also suggests limitations on the applicability in the pattern recognition sphere. However, we show that, with the enforcement of topological constraints on the reservoir, in particular that of small world topology, the model is indeed fault tolerant. Thus this implies that "natural" computational systems must have specific topologies and the uniform random connectivity is not appropriate.

Keywords: Reservoir Computing, Small world topology, robustness, Machine Learning.

1 Introduction

Recently Jaeger (Jaeger, "The écho state" approach to analysing and training recurrent neural networks, 2001), Maass (Maass, Natschläger, & Markram, 2002) and others have put forth the paradigm of "reservoir computing" as a way of computing with highly recurrent neural networks. This reservoir is a collection of neurons randomly connected with each other of fixed weights. Amongst other things, it has been shown to be effective in temporal pattern recognition; and has been held as a model appropriate to explain how certain aspects of the brain work. (Particularly in its guise as “liquid state machine”, due to Maass et al.)

This is particularly impressive as processing in artificial neurons typically is a-temporal. This is because the underlying basic neuronal model, that of McCullough-Pitts (McCullough & Pitts, 1943) is a-temporal by nature. As a result, most applications of artificial neural networks are related in one way or another to static pattern recognition. On the other hand, it has long been recognized in the brain science community that the McCullough-Pitts paradigm is inadequate. Various models of
differing complexity have been promulgated to explain the temporal capabilities (amongst other things) of natural neurons and neuronal networks.

However, during the last decade, computational scientists have begun to pay attention to this issue both from the neurocomputational and biological perspectives e.g. (Maass W., 1999; Maass, Natschläger, & Markram, 2002; Maass, Natschläger, & Markram, 2004; Fernando & Sojakka, 2005; Jaeger, The "echo state" approach to analysing and training recurrent neural networks, 2001), and investigations as to the computational capabilities of various models are being investigated.

Two such models, the “echo state machine” (Jaeger, The "echo state" approach to analysing and training recurrent neural networks, 2001) and the “Liquid State Machine” (see Fig. 1.) (Maass, Natschläger, & Markram, 2004; Maass, Natschläger, & Markram, 2002), have had substantial successes recently. These two models are identical on the abstract level and have recently been renamed “reservoir computing” (Lukosevicius & Jaeger, 2009). In these models there is a somewhat different paradigm of computation. Information is stored, not in "attractors" as is usually assumed in recurrent neural networks, but in the reverberating activity pattern in a sufficiently recurrent and inter-connected network. This information can then be retrieved by any sufficiently strong classifying detector. The idea is that the history of, e.g. timings of rocks thrown into a pond of water, is completely contained in the wave structure.) Moreover, the "persistence of the trace" (or as Maass put it, the "fading memory") allows one to recognize at a temporal distance the signal that was sent to the liquid; and sequence and timing affects of inputs.

This is an exciting idea; and, e.g. Jaeger, Maass and his colleagues have published a series of papers on it. Amongst other things, they have recently shown that once a detector has been sufficiently trained at any time frame, it is resilient to noise in the input data; and so it can be used successfully for generalization. (Maass, Natschläger, & Markram, 2002; Fernando & Sojakka, 2005). In particular, experiments have been performed for speech recognition.

However, there is a claim that this abstraction is faithful to the potential capabilities of the natural neurons and thus is explanatory to some extent from the viewpoint of computational brain science. It is this issue we address in this paper. Note that one of the underlying assumptions is that the detector works without memory; that is the detector should be able to classify based on instantaneous static information; i.e. by