Journal of Computer and System Sciences 82 (2016) 1232–1250

Contents lists available at ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

Expressive power of first-order recurrent neural networks determined by their attractor dynamics

Jérémie Cabessa [a,*], Alessandro E.P. Villa [b]

[a] Laboratory of Mathematical Economics (LEMMA), University Panthéon-Assas – Paris 2, 4 Rue Blaise Desgoffe, 75006 Paris, France
[b] Laboratory of Neuroheuristics (NHRG), Faculty of Business and Economics (HEC), University of Lausanne, CH-1015 Lausanne, Switzerland

A R T I C L E I N F O

Article history:
Received 19 October 2015
Received in revised form 31 March 2016
Accepted 19 April 2016
Available online 27 June 2016

Keywords:
Recurrent neural networks
Neural computation
Analog computation
Evolving systems
Learning
Attractors
Spatiotemporal patterns
Turing machines
Expressive power
ω-languages

A B S T R A C T

We provide a characterization of the expressive powers of several models of deterministic and nondeterministic first-order recurrent neural networks according to their attractor dynamics. The expressive power of neural nets is expressed as the topological complexity of their underlying neural ω-languages, and refers to the ability of the networks to perform more or less complicated classification tasks via the manifestation of specific attractor dynamics. In this context, we prove that most neural models under consideration are strictly more powerful than Muller Turing machines. These results provide new insights into the computational capabilities of recurrent neural networks.

1. Introduction

In the central nervous system, one neuron may receive and send projections from and to thousands of other neurons. The huge number of connections established by a single neuron and the slow integration time of neurons, operating in the milliseconds range (billion times slower than presently available supercomputers), suggest that information in the nervous system might be transmitted by simultaneous discharges of large sets of neurons. In addition, the presence of recurrent connections within large neural circuits indicates that re-entrant activity through chains of neurons should represent a major hallmark of brain circuits [5]. Also, developmental and/or learning processes are likely to potentiate or weaken certain pathways through the network by affecting the number or efficacy of synaptic interactions between the neurons [19,30]. Hence, the activation of functional cell assemblies in distributed networks might be induced by transmissions of complex patterns of activity [1]. In fact, various experimental studies suggest that specific attractor dynamics [20,21,56] as well as spatiotemporal pattern of discharges (i.e., ordered and precise interspike interval relationships) [2,49,51,53,54,57] are likely to be significantly involved in the processing and coding of information in the brain. Moreover, the association between attractor dynamics and repeating firing patterns has been demonstrated in nonlinear dynamical systems [3,4] and in simulations of large scale neuronal networks [28,29].

* Corresponding author.
E-mail address: jeremie.cabessa@u-paris2.fr (J. Cabessa).

http://dx.doi.org/10.1016/j.jcss.2016.04.006

On the other hand, since the late 1940's, the issue of the computational and dynamical capabilities of neural models has also been approached from a very theoretical standpoint [34]. In this context, neural networks are considered as abstract computing systems and their computational power is investigated from a theoretical computer scientist perspective [33–36, 39–42,48]. As a consequence, the computational power of neural networks has been shown to be intimately related to the nature of their synaptic weights and activation functions, and able to range from the finite state automata level [33–35] up to Turing [41] and super-Turing capabilities [8,10,39,40]. More recently, the Turing and super-Turing capabilities of various models of recurrent neural networks have been extended to the contexts of alternative bio-inspired paradigms of computation, like reactive-system-based computation [11–13,16] (i.e., abstract devices working over infinite input streams [38,46]) or interactive computation [6,9,15,17] (i.e., infinite sequential exchange of information between the system and its environment [27,50,60]).

Based on these theoretical and experimental considerations, we initiated the theoretical study of the expressive power of recurrent neural networks from the perspective of their attractor dynamics [16]. We proved that Boolean recurrent neural networks provided with two different kinds of attractors are computationally equivalent to Muller automata, and hence, recognize precisely the so-called ω-regular neural languages. Consequently, the most refined topological classification of ω-languages [59] can be transposed from the automaton to the neural network context, and yield to some transfinite hierarchical classification of Boolean neural networks according to their attractor dynamics [11,12]. This classification induces a novel attractor-based measure of complexity for Boolean recurrent neural networks, which notably refers to the ability of the networks to perform more or less complicated classification tasks, via the manifestation of meaningful or spurious attractor dynamics [16].

The present paper pursues this precise research direction and constitutes an extended version of the two proceedings papers [7,18]. Here, we provide a characterization of the expressive powers of various models of deterministic and nondeterministic sigmoidal (rather than Boolean) first-order recurrent neural networks, in terms of their attractor dynamics. This attractor-based expressive power also notably refers to the ability of the networks to perform more or less complicated classification tasks of their input streams via the manifestation of meaningful or spurious attractor dynamics, and hence, via the manifestation of meaningful or spurious spatiotemporal patterns of discharge.

The paper is organized as follows. In Section 2, we present the mathematical notions required for the development of our theory. In Section 3, we recall some basic definitions and facts concerning deterministic and nondeterministic Muller Turing machines.

In Section 4, the various models of deterministic and nondeterministic first-order recurrent neural networks under consideration are presented. These models are all based on classical first-order recurrent neural network composed with Boolean input and output cells as well as sigmoidal hidden units, along the very lines of [8,10,40,41]. The sigmoidal hidden neurons introduce the source of nonlinearity which is so important to neural computation. The Boolean input and output cells carry out the exchange of discrete information between the network and the environment. When subjected to some infinite binary input stream, the Boolean output cells necessarily exhibit some attractor dynamics, which is assumed to be of two possible kinds, namely either meaningful or spurious. The neural ω-language of a network then corresponds to the set of all input streams which induce a meaningful attractor dynamics. The expressive power of the networks is then measured via the topological complexity of their underlying neural ω-languages.

Section 5 provides the results of the paper, which are summarized in Subsection 5.1, and in particular, in Fig. 2 and Table 2. In short, the deterministic static rational-weighted neural networks are computationally equivalent to the deterministic Muller Turing machines, and every other model of deterministic or nondeterministic static or evolving neural nets is strictly more expressive than the Muller Turing machine model.

Subsection 5.2 concerns the expressive power of the deterministic neural networks. We show that the static rational-weighted and real-weighted neural networks are computationally equivalent to and strictly more powerful than the deterministic Muller Turing machines, respectively. Moreover, the evolving neural nets are equivalent to the static real-weighted ones, irrespective of whether their synaptic weights are modeled by rational or real numbers. They recognize precisely the set of all $BC(\mathbf{\Pi}^0_2)$ neural ω-languages (the finite Boolean combinations of $\mathbf{\Pi}^0_2$ neural ω-languages).

Subsection 5.3 focuses on the expressive power of the nondeterministic neural networks. Here, we introduce a novel learning-based notion of nondeterminism which encompasses the classical one studied by Siegelmann and Sontag [40, 41]. More precisely, we consider a model of evolving neural networks, where each network might a priori follow various patterns of evolution for its synaptic weights. At the beginning of each computation, the network selects one possible evolving pattern – in a nondeterministic manner – and then sticks to it throughout its whole computational process. We prove that the two models of rational-weighted and real-weighted nondeterministic neural networks are computationally equivalent, and recognize precisely the set of all $\mathbf{\Sigma}^1_1$ neural ω-languages. They are therefore strictly more expressive than the nondeterministic Muller Turing machines.

Finally, Section 6 discusses the interpretation and relevance of the results and provides some general concluding remarks.

2. Preliminaries

A *topological space* is a pair $(S, \mathcal{T})$ where S is a set and $\mathcal{T}$ is a collection of subsets of S such that $\emptyset \in \mathcal{T}$, $S \in \mathcal{T}$, and $\mathcal{T}$ is closed under arbitrary unions and finite intersections. The collection $\mathcal{T}$ is called a *topology* on S, and its members are called *open sets*. Given some topological space $(S, \mathcal{T})$, the class of *Borel subsets* of S, denoted by $\mathbf{\Delta}^1_1$, is the σ-algebra