Neural Networks 98 (2018) 318–336

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

A loop-based neural architecture for structured behavior encoding and decoding

Thomas Gisiger [a,*], Mounir Boukadoum [b]

[a] Centre for Research on Brain, Language and Music, 3640 de la Montagne, Montréal, Québec H3G 2A8, Canada
[b] Département d'informatique, Université du Québec à Montréal, Case postale 8888, succursale Centre-ville, Montréal Québec H3C 3P8, Canada

ARTICLE INFO

Article history:
Received 13 July 2017
Received in revised form 15 November 2017
Accepted 28 November 2017
Available online 8 December 2017

Keywords:
Neural networks
Functional integration
Gating
Basal ganglia
Artificial intelligence

ABSTRACT

We present a new type of artificial neural network that generalizes on anatomical and dynamical aspects of the mammal brain. Its main novelty lies in its topological structure which is built as an array of interacting elementary motifs shaped like loops. These loops come in various types and can implement functions such as gating, inhibitory or executive control, or encoding of task elements to name a few. Each loop features two sets of neurons and a control region, linked together by non-recurrent projections. The two neural sets do the bulk of the loop's computations while the control unit specifies the timing and the conditions under which the computations implemented by the loop are to be performed. By functionally linking many such loops together, a neural network is obtained that may perform complex cognitive computations. To demonstrate the potential offered by such a system, we present two neural network simulations. The first illustrates the structure and dynamics of a single loop implementing a simple gating mechanism. The second simulation shows how connecting four loops in series can produce neural activity patterns that are sufficient to pass a simplified delayed-response task. We also show that this network reproduces electrophysiological measurements gathered in various regions of the brain of monkeys performing similar tasks. We also demonstrate connections between this type of neural network and recurrent or long short-term memory network models, and suggest ways to generalize them for future artificial intelligence research.

1. Introduction

Artificial intelligence has recently made spectacular leaps forward in solving complex perception and prediction problems related to pattern recognition (*e.g., ImageNet competition*) and playing games at the level of expert human players (*e.g., AlphaGo of DeepMind*), to name a few successful achievements. Among the approaches and tools that made this possible, the deep learning paradigm for multi-layer feedforward neural networks (LeCun, Bengio, & Hinton, 2015) has probably had the most significant impact. Indeed, by adding a large number of hidden layers that are trained as unsupervised features detectors, deep learning substantially improves the classification performance of the standard supervised multi-layer perceptron with error backpropagation training (MLP-BP) architecture. Deep learning in its convolutional neural network form also brings some biological plausibility to the MLP model as a mimic of the main workings of the primary visual system of primates (Hubel & Wiesel, 1959, 1962).

As impressive as the advances brought forth by deep learning are, it remains to be shown whether it can scale up to more ambitious problems such as developing machines capable of self-awareness or a theory of mind (Fodor, 1981; Putnam, 1980). Taking a few steps back, one can argue that despite their prowess, neural networks equipped with deep learning currently fall short of exhibiting the rich dynamics and structural diversity exhibited by the brains of higher organisms at rest or when solving problems. Indeed, as has been shown using brain-imaging methods on subjects performing various tasks, the patterns of neural activity taking place in the brain vary considerably as a function of space, time and what the subjects are doing. First, the neural activation levels at different brain locations greatly vary in magnitude, with strong correlations with the behavior being currently executed (e.g., there is motor cortex activation when the subject is engaged in a motor task— Penfield & Boldrey, 1937); these variations are interpreted as the brain summoning the elementary functions (e.g., sensory or motor) necessary to perform the behavior at hand. Second, it was also found that the correlations between the neural activities in distinct brain regions change over time, especially for brain areas harboring functions useful to the current behavior (Friston et al., 1997). This is commonly interpreted as a higher-scale

* Corresponding author.
E-mail address: thomas.gisiger@mcgill.ca (T. Gisiger).
Abbreviations: CD, caudate nucleus; SNr, substantia nigra reticulata; GPi, internal segment of the globus pallidus

https://doi.org/10.1016/j.neunet.2017.11.019
0893-6080/

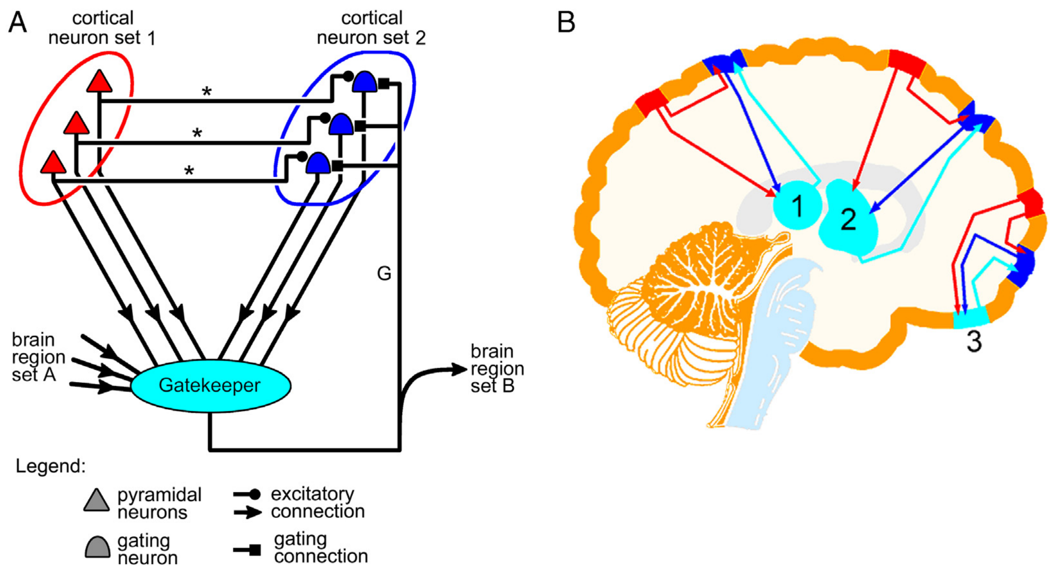

Fig. 1. Structure and location of gating loops. **A** General diagram of a gating loop spanning cortical neuron sets 1 and 2, and controlled by a third unit labeled gatekeeper. Information contained within neuron set 1 is carried over to neurons in set 2 by projections labeled with * depending on the state of the gatekeeper unit (modified from Gisiger and Boukadoum, 2011). **B** Diagram of three gating loops currently supported by experimental and theoretical evidence: loop controlled by neural circuitry located in the thalamus (1), the basal ganglia (2) or in cortex itself (3). (For interpretation of the references to color in this figure legend and how it encodes correspondance between structures in A and B, the reader is referred to the web version of this article.)
Source: See Gisiger and Boukadoum (2011) for details.

phenomenon where distinct brain regions become integrated together into larger functional entities as they exchange information with each other or collaborate to perform higher-level computations (Fuster, 1997). Further, the anatomical complexity of the brain's structure matches that of its dynamics; structural studies have shown that most neurons are localized in the cortex, where they are organized in local groups that are themselves interconnected in a complex network of longer-range reciprocal projections (Van Essen, Felleman, DeYoe, Olavarria, & Knierim, 1990).

Though those findings and interpretations seem reasonably well established, there is no clear consensus on the specific brain mechanisms underlying the observed dynamics when accomplishing given cognitive tasks. The options proposed in the literature encompass selection and competition (see Cocchi, Zalesky, Fornito, & Mattingley, 2013; Fuster, 1997 and references therein) and synchronization (Varela, Lachaux, Rodriguez, & Martinerie, 2001) to mention just a few. In this work, we present novel potential candidates for the mechanisms at play in a simple, yet neural-realistic form. This not only simplifies the grasping of the proposed concepts, but also opens the door for their use to implement some cognitive functions associated with the intelligence phenomenon as we will show. We base our discussion on the viewpoint that the neural subsets that implement simple functions in the brain behave as atomic functional entities, and that one should try to find how they interact with each other to generate the observed dynamics. Following this approach, Bressler (1995) suggested the existence of several types of essential interactions, one of them being gating. Gating refers to the control that a given neural population has over the information exchange taking place between two other neural groups (see Gisiger and Boukadoum, 2011 for a review and references therein).

Fig. 1**A** shows a general neural implementation of gating, where the information traveling from one group of neurons (group 1) to another (group 2) is either stopped at the input ("gate") of the destination group or granted passage depending on the state of neurons in a third "gatekeeper" population. Several mechanisms involving either inhibitory interneurons (Burchell, Faulkner, & Whittington, 1998; Olshausen, Anderson, & Van Essen, 1993, 1995; Vogels & Abbott, 2009) or bistable neurons (Grace, 2000; Kepecs & Raghavachari, 2007) have been proposed to account for the gating process, and work is still under way to better define them. Electrophysiological studies have also provided evidence that, depending on the cases considered, the gatekeeper circuitry might be located within the cortex or in subcortical areas of the brain (see Fig. 1**B** and Gisiger & Boukadoum, 2011 for a review).

To better understand how such gating mechanisms might fit in the overall brain circuitry, Gisiger and Boukadoum (2011) suggested a generic neural loop with gating as described in the previous paragraph and as shown in Fig. 1**A**, with the gate opening or closing depending on the activity of the gatekeeper unit, which can integrate information that is both intrinsic (activities of sets 1 and 2 in Fig. 1) and extrinsic (activity of other brain areas). Also, as we will see, communicating the gate state to other brain regions can have important and interesting computational properties (see also Gisiger and Boukadoum, 2011 for a discussion).

From a computational standpoint, gating mechanisms, with their ability to manage information communication between neural groups, allow them to either function independently (e.g., harbor unrelated activity patterns), or as one larger group to perform computations together. Such gating mechanisms have been proposed to solve delayed-response tasks where information must be sustained and protected in the mind of the subject over a delay in order to produce a response (Frank, Laughry, & O'Reilly, 2001; Gisiger & Kerszberg, 2006; Gisiger, Kerszberg, & Changeux, 2005). Gating mechanisms also allow the confining of synaptic plasticity and learning to connections controlled by gating units in their open state (Gisiger & Kerszberg, 2006, 2007; Gisiger et al., 2005).

Another essential feature of brain dynamics is inhibitory control (Fuster, 1997), which allows the brain to suppress excitatory activity using inhibition. Inhibitory control is often included in neural network simulations to suppress computations by-products, to eliminate activity coding for out-of-date information, or to weed out unproductive alternatives (see Gutkin, Laing, Colby, Chow, & Ermentrout, 2001 for an example). As will be argued here, it is likely that some form of inhibitory control, governed by neural loops and acting at the level of brain areas, exists in the brain for similar purposes.