Scalability properties of multimodular networks with dynamic gating

Dani Martí, Omri Barak, Mattia Rigotti, Stefano Fusi

Brain processes arise from the interaction of a vast number of elements. Despite the enormous number of participating elements, interactions are generally limited by physical constraints. Typically a neuron is connected to thousands other neurons, a far lower number than the hundred billion neurons in the brain. Unfortunately, it is the number of connections per neuron, not the total number of neurons, what often determines the performance of large neural networks (measured, e.g., as memory capacity), a fact that hinders the scalability of such systems.

We hypothesize that the scalability problem can be circumvented by using multimodular architectures, in which individual modules composed of local, densely connected recurrent networks interact with one another through sparse connections. We propose a general model of multimodular attractor neural networks in which each module state changes only upon external event and the change depends on the state of a few other modules. To implement this scheme, every module has to disregard the state of any module not involved in a particular interaction. Because a module can potentially interact with several others, ignoring the states of non-relevant modules would require learning of an exponentially large number of conditions.

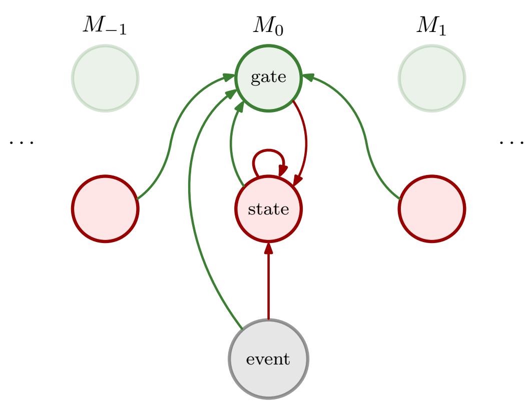
We solve this problem by adding a group of neurons that dynamically gate the interactions between modules. These neurons receive inputs from the modules and event signals through random sparse connections, and respond to combinations of event-states. This information is then sent back to the modules. Because they implement conjunctive representations, the number of necessary gating neurons grows only polynomially with the number of modules. We hypothesize that gating neurons reside in cortical layer 2/3, and that they mediate the interactions between modules in layer 5/6. The laminar organization of the neocortex could thus be a crucial architectural solution to the scalability problem.

Additional information

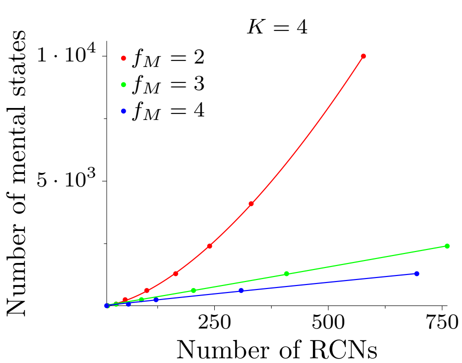
Imagine that we have M identical modules and that each module communicates to other $K < M$ modules, which we call the 'neighbors'. Each module can be in one of S possible states, encoded as stable patterns of activity. Changes in the configuration of the multimodular system are triggered by 'events', delivered as global external signals to all the modules. Let's denote the set of possible states by $\mathcal{S} = \{s_1, \ldots, s_S\}$, and the set of possible events by $\mathcal{E} = \{e_1, \ldots, e_E\}$. Our goal is to implement an arbitrary map $T : \mathcal{E} \times \mathcal{S} \times \mathcal{S}^K \to \mathcal{S}$ that specifies how the state of a module should transform upon the appearance of a particular event, given some particular configuration of neighbors. The number of combinations of state, events, and neighbor configurations is $S^{K+1}E$, which is potentially huge.

Because the number of constraints that can be simultaneously imposed scales linearly with the number of afferent connections (see, e.g., [1, 3]), there is no hope we can implement all the above rules in a system with limited connectivity. In realistic situations, however, each module needs to interact with only a few $f < K$ other modules at a time. And only a few states $S_{\text{eff}} < S$ of these other modules may actually induce transitions. In these cases, the number of relevant combinations of state, events, and neighbor configurations is reduced to S_{eff}^{f+1}. This number can easily drop to, or fall below, the number of afferent connections.

We propose a multimodular architecture that can implement such generic rule tables. The architecture incorporates a population of randomly connected neurons (RCNs), which are naturally selective to conjunctions of states and events [2]. With such mixed selectivity, these neurons can represent efficiently the rules to implement, and can route (i.e., dynamically gate) the interactions between modules. We show that the minimal number of RCNs necessary to implement random rule tables (with a given S, M, K, and f) depends polynomially on the number of modules, and is essentially dictated by the descriptive complexity of the rule table. Remarkably, the number of RCNs needed to implement a particular set of rules is only a few times bigger than the number of cells we would need in a carefully designed circuit [2].



Basic wiring diagram of a multimodular network with gated interactions The state of each module M_i is encoded by a population of recurrently connected neurons ('state' neurons). All state neurons receive inputs signaling external events ('event') as well as inputs from a population of gating neurons ('gate'). Importantly, the interaction between modules is not brought about by direct connections between 'state' neurons, but it is mediated by gating neurons. These are randomly innervated by state neurons from multiple modules, as well as by external neurons (green arrows: random sparse connections). Gating neurons transform nonlinearly this combination of inputs into firing rates, and this activity is fed back to the modules through plastic connections (shown as red arrows). These plastic connections can be trained off-line to drive all the set transitions described in a rule table.



Number of implementable mental states (or rules) as a function of the number of RCNs. Each module is connected to $K = 4$ neighbors. The parameter f_M is the number of modules each RCN receives inputs from. Each data point represents to the average number of RCNs needed to implement a sample of transition tables with a fixed S and f. The ordinate corresponds to the total number of entries that are compatible with the rules.

[1] D. O'Kane and A. Treves. Short- and long-range connections in autoassociative memory. *J. Phys. A-Math. Gen.*, 25(19):5055, 1992.

[2] M. Rigotti, D. B. D. Rubin, X. J. Wang, and S. Fusi. Internal representation of task rules by recurrent dynamics: the importance of the diversity of neural responses. *Front. Comput. Neurosci.*, 4, 2010.

[3] Y. Roudi and P. E. Latham. A balanced memory network. *PLoS Comput. Biol.*, 3(9):e141, 09 2007.