Darwinian evolution on neural networks: building in-silico systems to model cognitive processes

Research plan for four months

ANDRÁS SZILÁGYI

1. State of the art review of the research to date

The open-ended human problem solving (including e.g. complex thinking, insight problem solving or language acquisition) is far superior than what machine-learning can achieve. Despite many notable attempts, the underlying dynamics of complex thinking, e.g. insight problem solving, has not yet been understood. The Neuronal Replicator Hypothesis (NRH) argues that the essential key component of these processes is the dynamics of true Darwinian replicators within the brain – despite the fact that neurons do not reproduce. Why evolutionary dynamics? Evolutionary algorithms (EAs) are the "Swiss army knife" of mathematical methods in optimization and search processes. For certain problems special algorithms can overperfom an EA, but EA as a general multipurpose algorithm can solve *almost all* kinds of optimization/search problems. It is hard to imagine how a large set of domain-specific specialized algorithms is implemented within the brain. Thus we believe that real evolutionary processes might take place in the brain. Most of the previous attempts that combine Darwinian-like dynamics and cognitive processes/neural dynamics were sketchy or metaphorical e.g [1,2,3,4] or do not comply with evolutionary dynamics because no units of evolution (hereditary replicators, that can multiply) are postulated [5]. There are however promising attempts trying to couple neural dynamics with Darwinism. In these approaches, key components of Darwinian dynamics (variation, hereditary, fitness-evaluation, selection, reproduction) are present at different levels. In the last few years, the Neuronal Circuit Replicator theory [5], Neuronal Activity Replicator theory [6], Neuronal Classifier Replicator theory [7] and the Neuronal Path Replicator theory [8] formed a basis on which we can build our model, and made an interdisciplinary context in which entangling neuronal dynamics and evolution does not sound so unorthodox.

2. The importance of the research topic and how it relates to the thematic priorities of this call

Understanding the underlying large-scale dynamics of higher level complex thinking in humans is a very challenging task that raises many unanswered questions. One of the most amazing mental processes is insight problem solving. It occurs when a solution to a problem presents itself suddenly without gradient realization after many incorrect attempts based on trial and error. During insight the previous frame of reference of thinking is deconstructed and novel and more useful ("fitter") representations of a problem are created. It has become increasingly clear that genetic evolution is a process of insightful search [9], because it is also able to learn from past environments to structure and improve future search operators. These

remarkable and deep similarities justify the Neuronal Replicator Hypothesis that states that a Darwinian process of production of cognitive adaptations by natural selection can run in real-time in the neuronal network of the human brain. This project will provide the theoretical basis for neural replicator dynamics [10-13] of insight problem solution in line with the "Learning in evolutionary (social) systems characterized by replicator dynamics" project in the Cooperation and Conflict theme of the Institution (2016). As others have stated (pers.comm. Eörs Szathmáry) "*If we could show that Darwinian dynamics in the brain could help explain insightful search, this would be of no less importance than Darwin's contribution to the rest of biology.*"

3. The aim of the research

In line with the Neuronal Replicator Hypothesis, accepting the assumption that the key component of complex thinking and insight problem solving is a real Darwinian neurodynamics, I intend to work out the expected dynamics on the lowest possible level. This involves the investigation of the dynamics at the level of population of neural networks using mathematical models and computer simulations e.g. [14,15]. This research aims at integrating the analytical approach at the level of individual recurrent neural networks with simulations at the level of population of networks to form a solid theoretical and computational basis for the investigation of cognitive processes. The designed model system is expected to be independent of the human brain as medium, and whether Darwinian dynamics are truly present in the brain or not. Consequently, our model can be considered to be an effective standalone tool for search/optimization in high dimensional spaces using artificial neural networks architecture. The research is not only strongly synergistic with the proposed research plans of IZ and AF, but they are jointly depending on each other. Moreover, it is important to continuously maintain an intensive collaboration during model development between the three pillars of the NRH project: 1) theoretical background of Darwinian neurodynamics and neural network model development (A.S., this proposal), 2) evolutionary modeling and algorithmic formulation of real insight-problems (I.Z.) and human-subject cognitive experiments of insight problem solving (A.F.).

4. Innovative aspects of the research topic

Our target is to provide a biologically plausible basis for human problem solving and a bio-inspired mechanism for solving search and optimization problems. We would like to use the designed system as a tool specifically to analyze human insight problem solving. According to our knowledge this is the first approach dealing with true Darwinian neurodynamics of replicators not at the metaphorical level but mathematically firmly grounded. We implement the model on grounds of neurobiology, as a proof-on-principle, believing that real Darwinian neurodynamics offers a new, credible and efficient algorithm both for solving search/optimizing problems on neural networks and forming a solid theoretical and computational ground for the investigation of insight problem solving.

5. Detailed scheduled work plan for the grant period

1st month: Survey of the relevant literature of the theory of recurrent neural networks. Theoretical analysis of the memory capacity, investigation of the learning and forgetting processes. Measuring the pattern retrieval ability and the accurateness of retrieval. Investigation of the so-called "spurious patterns" as a potential source of intrinsic variability.

2nd month: Analysing the size and form of attractor basins in different network types. Computer implementation of the selected network(s). Analysis of the learning, forgetting and retrieval ability. Comparing the simulations and the theoretical results. Investigation of the effect of dilution of the connectivity matrix on the performance of the system. Measure the effect of storing similar patterns on the memory capacity and the correctness of retrieval.

3rd month: Extending the implementation to population of networks. Determining the maximum population size that can be handled to provide results in reasonable time by using workstations or servers. Defining a problem set with scalable difficulty (e.g. a fitness landscape with variable dimensionality and correlatedness). Determining the relevant parameters of the model (population size, mutation rate, network size, etc.) for effective evolutionary search on the above-defined problem set.

4th month: Using the developed modeling architecture to solve different kinds of problems. Testing the system as a bio-inspired optimization method running on neural networks and applying it to simplified model problems related to insight problem solving. Combining dependent and complementary results of the other two parts of the investigation (I.Z. and A.F). Preparing a publication to disseminate results.

6. Planned work methodology

The proposed investigation is based on two different methodological aspects. The first is theoretical, based on extensive mathematical analysis of different kinds of neural networks; the second is a large-scale *in-silico* analysis of populations of these networks. The first includes the following items: 1) formalization of different kinds of recurrent neural networks (RNNs), analyzing the memory capacity and retrieval ability; 2) theoretical analysis of the learning and forgetting processes to avoid “catastrophic forgetting” and implementing palimpsest-type memory [16]; 3) investigation of the dynamical stability by surveying the relevant theoretical literature; 4) further analyzing and extending present theoretical models of RNNs. The second approach is computational. After finding a suitable model(s) of RNNs for numerical investigation we will design a computer implementation. Because large populations of networks are needed for the Darwinian dynamics, the programming will be made in low level languages (preferably in C). Small scale simulations (with a single network) is necessary to test the behavior of chosen recurrent neural network model (memory, learning, forgetting, etc.) then large scale simulations on population of RNNs to analyze the possibility and effectiveness of the Darwinian dynamics on this architecture. Using problems of increasing complexity and dimensionality we will analyze the computational capacity needed to perform

large-scale investigations (detailed in the proposals of I.Z. and A.F.). We plan to collaborate with the Big Data program by using its computational capacities.

7. Expected research results and their practical utilization

At the end of the grant period, a working computer model of populations of recurrent neural networks is expected. This model framework opens up the way of using it at least in two contexts. Firstly it can be used as an optimization algorithm implemented on neural network architecture (or "hardware"). This applicability is independent of the real dynamics present in the human brain and the procedure can be a very effective optimization and search method especially in high dimensional problem spaces. Secondly, proving the assumption that real Darwinian dynamics can operate on neural networks forms a solid basis for further investigation of complex thinking, cognitive processes and insight problem solving joining the two other proposal submitted in this topic. The synergistic interaction of the three researches are expected to form a complete research agenda, providing a proof of principle through different levels: from individual neurons through populations of neurons and neuronal networks to a baseline, simplified model of human insight problem solving.

Requirements

For successful completion of the proposed research plan accessing to international journals is necessary. For the theoretical analysis and model development, moderate computation capacity (desktop computers) is enough, but for large-scale simulations I would like to have access to the servers of the Big Data center. I have been working in this field in the last 10 years, acquiring a broad understanding in many fields of modeling biological systems, including neural networks with multiple publications in high quality international journals. I am confident that I am more than capable of achieving the goals of this proposal, though I would like to ask you to consider providing a grant between the standard post-doctoral and senior researcher levels.

References

[1] Bateson, G. (1979) *Mind and Nature: A Necessary Unity.* Bantam Books.

[2] Changeux, J. P. (1985) *Neuronal Man: The Biology of Mind.* Princeton University Press.

[3] Edelman, G. M. (1987) *Neural Darwinism. The Theory of Neuronal Group Selection*. New York, Basic Books.

[4] Price, G.R., (1995) *The Nature of Selectio.n* Journal of Theoretical Biology, 175(3): p. 389-396.

[5] Fernando, C., Karishma, K. K., & Szathmáry, E. (2008) *Copying and Evolution of Neuronal Topology.* PLoS ONE, 3(11): p. e3775.

[6] Fernando, C., Goldstein, R., et al., (2010). *The Neuronal Replicator Hypothesis.* Neural Computation 22(11): 2809–2857.

[7] Fernando, C. (2011) *Symbol Manipulation and Rule Learning in Spiking Neuronal Networks.* Journal of Theoretical Biology 275: 29-41 doi:10.1016/j.jtbi.2011.01.009

[8] Fernando, C., Vasas, V., et al. (2011) *Natural Selection of Paths in Networks.* Available from Nature Precedings http://hdl.handle.net/10101/npre.2011.5535.1 (2011)

[9] Zylberberg, A., Fernández Slezak, D., Roelfsema, P.R., Dehaene, S., Sigman, M. (2010*) The brain's router: A cortical network model of serial processing in the primate brain*. PLoS Computational Biology 6(4):e1000765.

[10] Maynard Smith, J. (1987*) How to model evolution*, ed. Dupré J. (MIT Press, Cambridge, MA), pp. 117-131.
[11] Harper, M. (2009) *The replicator equation as an inference dynami.c* ArXiv e-prints.
[12] Koza, J.R. (1992*) Genetic programming: On the programming of computers by natural selection*. (MIT Press, Cambridge, MA), p. 819.
[13] Fernando, C., Szathmáry, E. (2010*) Natural selection in the brain in Towards a theory of thinking*, On thinking, eds. Glatzeder B, Goel V, von Müller A. (Springer-Verlag, Berlin/Heidelberg) Vol. 5, pp. 291–322.
[14] Hopfield, J.J. (1982) *Neural networks and physical systems with emergent collective computational abilities.* Proceedings of the National Academy of Sciences 79(8):2554–2558.
[15] Rolls, E.T., Treves, A. (1998) *Neural networks and brain function*. (Oxford University Press, Oxford, New York).
[16] Storkey, A.J. (1998) *Palimpsest memories: a new high-capacity forgetful learning rule for Hopfield networks*. Technical report.