Is neocortical encoding of sensory information intelligent?

András Lőrincz

Eötvös Loránd University, 1117 Budapest, Pázmány Péter sétány 1/D, Hungary

Abstract

The theory of computational complexity is used to underpin a recent model of neocortical sensory processing. It is argued that the theory of computational complexity points to generative networks and that these networks resolve the homunculus fallacy. Computational simulations illustrate the idea.

Key words: computational complexity, neocortical processing, homunculus fallacy, reconstruction network

1 Introduction

A recent model of neocortical information processing developed a hierarchy of reconstruction networks subject to local constraints [1]. Mapping to the entorhinal-hippocampal loop has been worked out in details [2]. Straightforward but falsifying predictions of the model concern the temporal properties of the internal representation and the counteraction of delays of the reconstruction process. These predictions have gained independent experimental support recently [3,4]. The contribution of the present work is to underpin the model by the theory of computational complexity (TCC) and to use TCC to ground the concept of intelligence (CoI). We shall argue that the same considerations resolve the homunculus fallacy [5].

2 Theoretical considerations

In our view, the problem of encoding information in the neocortical sensory processing areas may not be detachable from CoI. We assume that the wiring of neocortical sensory processing areas developed by evolution forms an ensemble of economical intelligent agents and we pose the question: What needs to be communicated between intelligent computational agents? The intriguing issue is that although (i) CoI has meaning for us and (ii) this meaning seems to be measurable in practice, nevertheless, (iii) CoI has escaped mathematical definition. In turn, our task is twofold: we are to provide a model of neocortical processing of sensory information *and* a computational definition of intelligence.

According to one view , intelligent agents learn by developing categories [6]. For example, mushroom-categories could be learned in two different ways: (1) by 'sensorimotor toil', that is, by trial-and-error learning with feedback from the consequences of errors, or (2) by communication, called 'linguistic theft', that is, by learning from overhearing the category described. Our point is that case (2) requires mental verification: Without mental verification trial-by-error learning is still a necessity. In our model, verification shall play a central role for constructing the subsystems, our agents.

Problem solving and verification can be related by TCC. From the point of view of communication, there are only two basic types of computational tasks. The first, which can be called as not worth to communicate (non-WTC) type is either easy to solve and easy to verify, or hard to solve and hard to verify. The other type is hard to solve but easy to verify and, in turn, it is of WTC type. For non-WTC type problems communication is simply an overhead and as such, it is not economical. On the other hand, WTC type problems – according to TCC – may have exponential gains if communication and then verification is possible. As an example, consider the NP-hard Traveling Salesman problem. The complexity of the problem is known to scale exponentially with the number of cities, whereas time of verification scales linearly. As another problem, consider 'recognition by components'. This is also a network problem and it is similar to TSP: The word city needs to be replaced by the word component.

We conclude that economical communication occurs only in WTC-type problems. The intelligent 'agents' (the subsystems) are subject to the following constraints: (1) Subsystems need to learn to separate components of a combinatorial tasks. (2) Information provided by the subsystems are to be joined, e.g., at a higher level. To highlight the concept, the following definition is constructed.

Definition. We say that an intelligent system is embodied in an intelligent, possibly hierarchical environment, if (a) it can learn and solve combinatorial problems, (b) can communicate the solutions, and (c) if it can be engaged in (distributed) verification of solutions communicated by other intelligent agents.

To build a network model, verification is identified as the opposite of encoding. Verification of an encoded quantity means (i) decoding, i.e., the reconstruction of inputs using communicated encoded quantities, (ii) comparison of the reconstructed input and the real input. In turn, a top-down model of neocortical processing of sensory information can make use of generative models equipped with comparators, in which the distributed hierarchical decoding process is to be reinforced by comparisons. This is our computational model for CoI.

3 Results and discussion

The reconstruction network model. Encoding, decoding and comparison is performed by reconstruction networks. The basic reconstruction network (Fig. 1A) has two layers: the reconstruction error layer that computes the difference ($\mathbf{e} \in \mathrm{R}^r$) between input ($\mathbf{x} \in \mathrm{R}^r$) and reconstructed input ($\mathbf{y} \in \mathrm{R}^r$): $\mathbf{e} = \mathbf{x} - \mathbf{y}$ and the hidden layer that holds the hidden representation $\mathbf{h} \in \mathrm{R}^s$ and produces the reconstructed input $\mathbf{y}$ via top-down transformation $\mathbf{Q} \in \mathrm{R}^{r \times s}$. The hidden representation is corrected by the bottom-up transformed form of the reconstruction error $\mathbf{e}$, i.e., by $\mathbf{We}$, where $\mathbf{W} \in \mathrm{R}^{s \times r}$ and is of rank $\min(s, r)$. The process of correction means that the previous value of the hidden representation is to be maintained and the correcting amount needs to be added. In turn, the hidden representation has self-excitatory connections ($\mathbf{M}$), which sustains the activities. For sustained input $\mathbf{x}$, the iteration will stop when $\mathbf{WQh} = \mathbf{Wx}$: The relaxed hidden representation is solely determined by the input and top-down matrix $\mathbf{Q}$. The latter is identified with the long-term memory. BU matrix $\mathbf{W}$ is perfectly tuned if $\mathbf{W} = (\mathbf{Q}^T\mathbf{Q})^{-1}\mathbf{Q}^T$, i.e., if $\mathbf{WQ} = \mathbf{I}$ ($\mathbf{I} \in \mathrm{R}^{s \times s}$). In this case, the network is as fast as a feedforward net.

The network can be extended to support noise filtering by (i) separating the reconstruction error layer and the reconstructed input layer, (ii) adding another extra layer that holds $\mathbf{s} = \mathbf{We}$ and transformation $\mathbf{s} \to \mathbf{h}$, i.e., $\mathbf{N}$ supports pattern completion [1] and (iii) assuming that BU transformation $\mathbf{W}$ maximizes BU information transfer by minimizing mutual information (MMI) between its components. BU transformed error is passed to the hidden representation layer through transformation matrix $\mathbf{N}$ and corrects hidden representation $\mathbf{h}$. Hidden representation may make use of e.g., non-negative matrix factorization [7] (Fig. 1(B)). MMI plays an important role in noise filtering. There are two different sets of afferents to the MMI layer: one carries the error, whereas the other carries the reconstructed input $\mathbf{y}$ via bottom-up transformation $\mathbf{P}$, followed by a non-linearity that removes noise via thresholding. Thresholding is alike to wavelet denoising, but filters are not necessarily wavelets: They are optimized for the input database experienced by the network. MMI algorithms enable the local estimation and local thresholding of noise components. The method is called sparse code shrinkage (SCS) [8]. Note that SCS concerns the

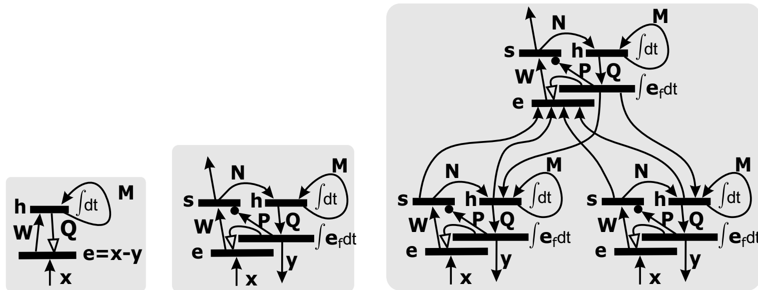

Fig. 1. **Reconstruction networks**
A: Simple reconstruction network (RCN). **B:** RCN with sparse code shrinkage noise filtering. **C:** RCN hierarchy. (See text.)

components of the BU transformed reconstructed input: high (low) amplitude components of the BU transformed reconstructed input $\mathbf{Py}$ can (not) open the gates of components of the MMI layer and MMI transformed reconstruction error can (not) pass the open (closed) gates to correct the hidden representation. Apart from SCS, the reconstruction network is linear. We shall denote this property by the sign '$\sim$'. '$A \sim B$' means that up to a scaling matrix, quantity A is approximately equal to quantity B. For a *well tuned network* and if matrix $\mathbf{M}$ performs temporal integration, then $\dot{\mathbf{x}} \cong \mathbf{e} \sim \mathbf{s}$ by construction. In a similar vein, hidden representation $\mathbf{h} \sim \mathbf{y}$, apart from noise, $\mathbf{y} \sim \mathbf{x}$, and $\mathbf{y} = \int \mathbf{e}_f dt$ where $\mathbf{e}_f$ is the noise filtered version of $\mathbf{e}$. In a hierarchy of such networks, e.g., $\mathbf{h}$ or $\dot{\mathbf{x}}$ or both can be transmitted to higher networks. Higher networks may overwrite the internal representation of lower networks. Non-linear effects are introduced by SCS and by the top-down overwriting of internal representation. (Fig. 1C).

Resolving the homunculus fallacy. The generative network concept provides a straightforward explanation [9] to the homunculus fallacy (see, e.g., [5]). The homunculus fallacy says that the internal representation is meaningless without an interpreter and all levels of abstraction require at least one further level to become the corresponding interpreter. Unfortunately, the interpretation — according to the fallacy — is just a new transformation and we are trapped in an endless regression. Our standpoint is that the paradox stems from vaguely described procedure of 'making sense'. The fallacy arises by saying that the internal representation should make sense. One can turn the fallacy upside down by changing the roles[9]: Not the internal representation but the *input* should make sense. Our proposal is that the *input makes sense* if the same (or similar) inputs have been experienced before and if the input can be derived or regenerated by means of the internal representation [1]. Reconstruction networks constrain the infinite regression into iterative dynamics within a finite architecture and, in turn, the fallacy disappears. According to this approach the internal representation interprets the input by (re)constructing it.

4 Computational simulations

A combinatorial problem was studied in [1]. In these simulations, several horizontal and vertical bars were presented to the network similar to the one depicted in Fig. 1C. Subnetworks had local receptive fields and did separate the components, the horizontal and vertical bars in the internal representation. The architecture joined bottom-up MMI noise filtering and top-down non-negative matrix factorization (NMF) [7]. The reconstruction network higher in the hierarchy collected information from the lower networks and overwrote the internal representation. Encoded information collected from and communicated to lower networks enabled improved pattern completion. For illustrative purposes, we review the results here:

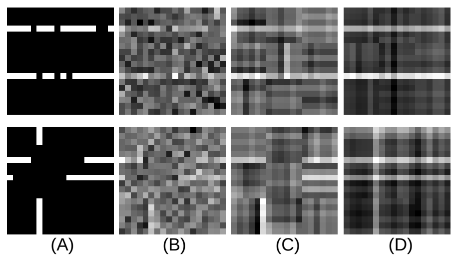

Fig. 2. **Improved noise filtering and pattern completion in the hierarchy.** Sub-architectures of the first layer have $6 \times 6 = 36$ input dimension. Dimension of NMF hidden vectors of first layer units is 12. First layer is made of $3 \times 3 = 9$ units. Input of the second layer has 108 dimensions. Dimension of the hidden vector of the second layer is 36. *Upper row:* Pixels of the inputs are missing. *Bottom row:* Pixels *and* sub-components of the inputs are missing. (A) Original input with missing pixels and sub-components, (B) noise covered input received by the architecture (during learning, too), (C) reconstructed input using first layer reconstructions only, (D) reconstructed input using the full hierarchy.

5 Conclusions

Our goal was to underpin a recent model of neocortical information processing [1] by means of the theory of computational complexity. We have argued that sensory processing areas developed by evolution should be viewed as intelligent agents subject to economical constraints on communication. According to our argument, the agents encode solutions to combinatorial problems (NP-hard problems in general, or 'components' according to psychology [10]), communicate the encoded information and decode the communicated information. A computational model of 'linguistic theft' can be built by considering that mushroom categories are built from combinations of features, such as combinations of basic geometric forms, coloring of the different parts, environment, etc. We have reviewed computational experiments where a combinatorial component search problem was (1) solved/learned and (2) the communicated encoded information was decoded and used to improve pattern completion. The novelty of the present work is in the interpretation of previous results, that is the computational model of intelligence; a set of agents that (i) communicate encoded information about combinatorial learning tasks and (ii) cooperate in input verification via decoding of the communicated encoded quantities. We have shown that the same model offers a straightforward solution for the homunculus fallacy.

Acknowledgements. This work was supported by OTKA (No. T-32487).

References

[1] A. Lőrincz, B. Szatmáry, G. Szirtes, Mystery of structure and function of sensory processing areas of the neocortex: A resolution, J. Comp. Neurosci. 13 (2002) 187–205.

[2] A. Lőrincz, G. Buzsáki, in: Annals of the New York Acad. Sci., Vol. 911, New York, 2000, Ch. Two–phase computational model training long–term memories in the entorhinal–hippocampal region, pp. 83–111.

[3] A. V. Egorov, B. N. Hamam, E. Fransén, M. E. Hasselmo, A. A. Alonso, Graded persistent activity in entorhinal cortex neurons, Nature 420 (2002) 173–178.

[4] D. A. Henze, L. Wittner, G. Buzsáki, Single granule cells reliably discharge targets in the hippocampal ca3 network in vivo, Nature Neurosci. 5 (2002) 790–795.

[5] J. R. Searle, The rediscovery of mind, Bradford Books, MIT Press, Cambridge, MA, 1992.

[6] S. Harnad, Can a machine be conscious? How?, http://cogprints.ecs.soton.ac.uk/archive/00002460/ (2003).

[7] D. D. Lee, H. S. Seung, Algorithms for non-negative matrix factorization, in: Advances in Neural Processing Systems, Vol. 13, Morgan Kaufmann, San Mateo, CA, 2001, pp. 556–562.

[8] A. Hyvärinen, Sparse code shrinkage: Denoising of nongaussian data by maximum likelihood estimation, Neural Computation 11 (1999) 1739–1768.

[9] A. Lőrincz, Towards a unified model of cortical computation II: From control architecture to a model of consiousness, Neural Network World 7 (1997) 137–152.

[10] I. Biederman, Recognition-by-components: a theory of human image understanding, Psychol. Rev. 94 (1987) 115–147.