Towards Developmental AI: The paradox of ravenous intelligent agents

Michelangelo Diligenti, Marco Gori, and Marco Maggini
DII - University of Siena

1 Introduction

In spite of extraordinary achievements in specific tasks, nowadays intelligent agents are still striving for acquiring a truly ability to deal with many challenging human cognitive processes, especially when a mutable environment is involved. In the last few years, the progressive awareness on that critical issue has led to develop interesting bridging mechanisms between symbolic and sub-symbolic representations and to develop new theories to reduce the huge gap between most approaches to learning and reasoning. While the search for such a unified view of intelligent processes might still be an obliged path to follow in the years to come, in this paper, we claim that we are still trapped in the insidious paradox that feeding the agent with the available information, all at once, might be a major reason of failure when aspiring to achieve human-like cognitive capabilities. We claim that the children developmental path, as well as that of primates, mammals, and of most animals might not be primarily the outcome of biologic laws, but that it could be instead the consequence of a more general complexity principle, according to which the environmental information must properly be filtered out so as to focus attention on "easy tasks." We claim that this leads necessarily to stage-based developmental strategies that any intelligent agent must follow, regardless of its body.

2 Developmental path and focus of attention

There a number of converging indications that most of nowadays approaches to learning and reasoning have been bouncing against the same wall (in the case of sequential information, see e.g. [Frasconi *et al.*, 1995]). This is especially clear when facing cognitive tasks that involve both learning and reasoning capabilities, that is when symbolic and sub-symbolic representations of the environment need to be properly bridged. A unified approach to embrace the behavior of intelligent agents involved in both perceptual and symbolic information is based on expressing learning data and explicit knowledge by constraints [Diligenti *et al.*, 2010]. Following that framework, let us consider tasks that can be formalized by expressing a *parsimonious* solution consistent with a given set of constraints $\mathcal{C} = \{\chi_1, \ldots, \chi_q\}$. It is worth mentioning that the in the case of supervised learning, the parsimonious satisfaction of the constraints is reduced to a finite collection of points according to the classic statistical framework behind kernel machines. We consider an agent which operates dynamically in a mutable environment where, at each time t, it is expected to access only a limited subset $\mathcal{C}_t \subset \mathcal{C}_U$ of constraints, where $\mathcal{C}_U$ can be thought of as the *universal set of constraints*. Of course, any agent of relevant interest might be restricted to acquire a limited set of constraints $\mathcal{C}$, so as $\forall t \in \mathcal{T} : \ \mathcal{C}_t \subset \mathcal{C} \subset \mathcal{C}_U$. Instead of following a *developmental path*, one could think of agents that acquire $\mathcal{C}$ all at once.

Definition 2.1 *A ravenous agent is one which accesses the whole constraint set at any step, that is one for which* $\forall t \in \mathcal{T} : \ \mathcal{C}_t = \mathcal{C}$.

At first a glance, ravenous agents seem to have more chances to develop an efficient and effective behavior, since they can access all the information expressed by $\mathcal{C}$ at any time. However, when bridging symbolic and sub-symbolic models one often faces the problem of choosing a developmental path. It turns out that accessing all the information at once might not be a sound choice in terms of complexity issues.

The paradox of ravenous agents: *Ravenous agents are not the most efficient choice to achieve a parsimonious constraint consistency.*

To support the paradox, we start noting that hierarchical modular architectures used in challenging perceptual tasks like vision and speech are just a way to introduce intermediate levels of representation, so as to focus on simplified tasks. For example, in speech understanding, phonemes and words could be intermediate steps for understanding and take decisions accordingly. Similarly, in vision, SIFT features could be an intermediate representation to achieve the ability to recognize objects. However, when looking for deep integration of sub-symbolic and symbolic levels the issue is more involved and mostly open. We discuss three different contests that involve different degree of symbolic and sub-symbolic representations.

Developmental paths

- *Reasoning in the environment* - When thinking of circumscription and, in general, of a non-monotonic reasoning, one immediately realizes that we are addressing issues that are outside the perimeter of ravenous agents. For example, we can start from a default assumption, that is the typical "bird flies." This leads us to conclude that if a given animal is known to be a bird, and nothing else is known, it can be assumed to be able to fly. The default assumption can be retracted in case we subsequently learn that the animal is a penguin. Something similar happens during the learning of the past tense of English verbs [Rumelhart and McClelland, 1986], that is characterized by three stages: memorization of the past tense of a few verbs, application of the rule of regular verbs to all verbs and, finally, acquisition of the exceptions. Related mechanisms of retracting previous hypotheses arise when performing abductive reasoning. For example, the most likely explanation when we see wet grass is that it rained. This hypothesis, however, must be retracted if we get to know that the real cause of the wet grass was simply a sprinkler. Again, we are in presence of non-monotonic reasoning. Likewise, if a logic takes into account the handling of something which is not known, it should not be monotonic. A logic for reasoning about knowledge is the autoepistemic logic, which offers a formal context for the representation and reasoning of knowledge about knowledge. Once again, we rely on the assumption of not to construct ravenous agents, which try to grasp all the information at once, but on the opposite, we assume that the agent starts reasoning with a limited amount of information on the environment, and that there is a mechanism for growing up additional granules of knowledge.

- *Nature helps developing vision* - The visual behavior of some animals seems to indicate that motion plays a crucial role in the process of scene understanding. Like other animals, frogs could starve to death if given only a bowlful of dead flies [1], whereas they catch and eat moving flies [Lettvin *et al.*, 1968]. This suggests that their excellent hunting capabilities depend on the acquisition of a very good vision of moving objects only. Saccadic eye movements play an important role in facilitating human vision and, in addition, the autonomous motion is of crucial importance for any animal in vision development. Birds, some of which exhibit proverbial abilities to discover their preys (e.g. eagles, terns), are known to detect slowly moving objects, an ability that is likely to have been developed during evolution, since the objects they typically see are far away when flying. A detailed investigations on fixational eye movements across vertebrate indicates that micro-saccades appear to be more important in foveate than afoveate species and that saccadic eye movements seem to play an important role in perceiving static images [Martinez-Conde and Macknik, 2008]. When compared with other animals, humans are likely to perform better in static - or nearly static - vision simply because they soon need to look at objects and pictures thoughtfully. However, this comes at a late stage during child development, jointly with the emergence of other symbolic abilities. The same is likely to hold for amodal perception and for the development of strange perceptive behavior like popular Kanizsa triangle [Kanizsa, 1955]. We claim that image understanding is very hard to attack, and that the fact that humans brilliantly solve the problem might be mostly due to the natural embedding of pictures into visual scenes. Human perception of static images seems to be a higher level quality that might have been acquired only after having gained the ability of detecting moving objects. In addition, the saccadic eye movements might suggest that static images are just an illusory perception, since human eyes always perceive moving objects. As a consequence, segmentation and recognition are only apparently separate processes: They could be in fact two faces of the same medal that are only regarded as separate phases mostly because of the bias induced by years of research in pattern recognition aimed at facing specific perceptual tasks. Interestingly, animals which develop a remarkable vision system, and especially humans, in their early stages characterized by scarce motion abilities, deal primarily with moving objects from a fixed background, which facilitates their segmentation and, consequently, their recognition. Static and nearly static vision comes later in human developments, and it does not arise at all in many animals like frogs. The vision mechanisms seems to be the outcome of complex evolutive paths (e.g. our spatial reasoning inferences, which significantly improve vision skills, only emerge in late stage of development). We subscribe claims from developmental psychology according to which, like for other cognitive skills, its acquisition follows rigorous stage-based schemes [Piaget, 1961]. The motion is likely to be the secret of vision developmental plans: The focus on quickly moving objects at early stages allows the agent to ignore complex background.

- *On the bridge between learning and logic*
  Let us consider the learning task sketched in Fig. 2, where supervised examples and FOL predicates can be expressed in the same formalism of constraints. There is experimental evidence to claim that a ravenous agent which makes use of all the constraints $\mathcal{C}$ (supervised pairs and FOL predicates) is not as effective as one which focuses attention on the supervised examples and, later on, continues incorporating the predicates [Diligenti *et al.*, 2010]. Basically, the developmental path which first favors the sub-symbolic representations leads to a more effective solution. The effect of this developmental plan is to break up the complexity of learning jointly examples and predicates. Learning turns out to be converted into an optimization problem, which is typically plagued by the presence of sub-optimal solutions. The developmental path which enforces the learning from examples at the first stage is essentially a way to circumvent local minima.

[1] In addition to their infrared vision, snakes are also known to react much better to quick movements.

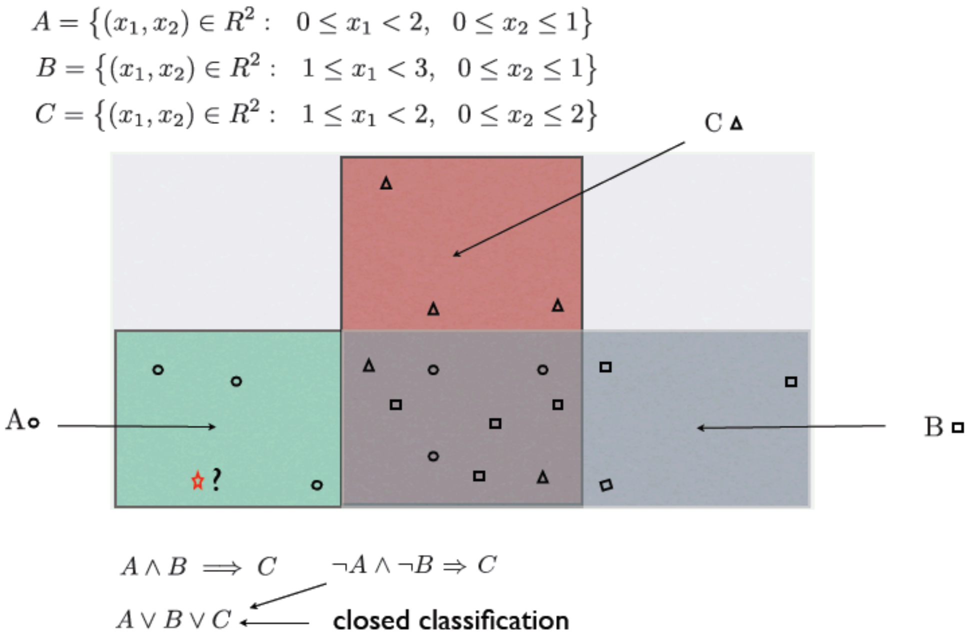

Figure 1: The task consists of learning three classes from examples and from a set of FOL predicates. The ordering of presentation does matter.

3 Conclusions

This paper supports the position that stage-based learning, as discussed in developmental psychology is not the outcome of biology, but is instead the consequence of optimization principles and complexity issues that hold regardless of the body. This position is supposed to re-enforce recent studies on developmental AI more inspired to studies in cognitive development (see e.g. [Sloman, 2009]) and is somehow coherent with the growing interest in deep architectures and learning [Bengio *et al.*, 2009].

Acknowledgements: We thank Luciano Serafini for insightful discussions.

References

[Bengio *et al.*, 2009] Y. Bengio, J. Louradour, R. Lollobert, and J. Weston. Curriculum learning. In *26th Annual International Conference on Machine Learning*, pages 41–48, Montreal, Canada, June 2009.

[Diligenti *et al.*, 2010] M. Diligenti, M. Gori, M. Maggini, and L. Rigutini. Multitask kernel-based learning with logic constraints. In *The 19th European Conference on Artificial Intelligence*, 2010.

[Frasconi *et al.*, 1995] P. Frasconi, M. Gori, and G. Soda. Recurrent neural networks and prior knowledge for sequence processing: a constrained non-deterministic approach. *Knowledge-Based Systems*, 8(6):313–332, 1995.

[Kanizsa, 1955] G. Kanizsa. Margini quasi-percettivi in campi con stimolazione omogenea. *Rivista di Psicologia*, 49(1):7–30, 1955.

[Lettvin *et al.*, 1968] J.Y. Lettvin, H.H. Maturana, W.S. McCulloch, and W.H. Pitts. What the frog's eye tells the frog's brain. In *Reprinted from - The Mind: Biological Approaches to its functions - Eds Willian C. Corning, Martin Balaban*, pages 233–358. 1968.

[Martinez-Conde and Macknik, 2008] S. Martinez-Conde and S.L. Macknik. Fixational eye movements across vertebrates: Comparative dynamics, physiology, and perception. *Journal of Vision*, 8(14):1–16, 2008.

[Piaget, 1961] J. Piaget. *La psychologie de l'intelligence*. Armand Colin, Paris, 1961.

[Rumelhart and McClelland, 1986] D.E. Rumelhart and J.L. McClelland. On learing the pat tense of english verbs. In *Parallel Distributed Processing, Vol. 2*, pages 216–271. 1986.

[Sloman, 2009] A Sloman. Ontologies for baby animals and robots. from baby stuff to the world of adult science: Developmental ai from a kantian viewpoint. Technical report, University of Birmingham, 2009.