Semantic-based regularization and Piaget's cognitive stages [*]

Marco Gori
University of Siena

April 1, 2009

ARE COGNITIVE STEPS A SECRET TO TO BREAK THE COMPLEXITY OF LEARNING?

In the last decades, the focus on the research in machine learning has hovered pendulum-like from biologically inspired to artificial models with different degree of cognitive plausibility. As the time goes by and research evolves, we will likely see men and machines facing an increasingly number of similar learning tasks. Now, regardless of the extent to which they share biological principles, one might be interested in studying human and artificial cognitive processes under the same umbrella. In this paper we claim that the principles of cognitive development stages, that have been the subject of an in-depth analysis in children by Jean Piaget [1, 2] are likely to inspire important advances in machine learning. He pointed out that we can identify four major stages or periods of development in child learning, where each stage is self-contained and builds upon the preceding stage. In addition, children seem to proceed through these stages in a universal, fixed order. They start developing sensorimotor and preoperational skills, in which the perceptual interactions with the environment dominate the learning process, and evolve by exhibiting concrete and formal operational skills, in which they start to think logically and develop abstract thoughts. When observing human and nowadays artificial minds on the same play, one early realizes that machines do not take into account most of the rich human communication protocols. In most of the studies of machine learning, the agent is expected to learn from labelled and/or un-labelled examples finalized to a specific task. There are, however, a number of other crucial interactions of the agent that are rarely taken into account. Human learning experiences witness the importance of asking questions and of learning under a of teaching plan. While the first interaction has been considered in a number of machine learning models, apart from a remarkable exception [3], to the best of our knowledge, teaching plans have not been significantly involved in learning algorithms. What is often neglected in machine learning is that most intriguing human learning skills are due, to a large extent, to the acquisition of relevant semantic attributes and to their relations. This makes learning a process which goes well beyond pure induction; the evidence provided by the induction of a semantic attribute is typically propagated to other attributes by formal rules, thus giving rise to a sort of reinforcement cyclic process.

[*]When I was in high school, my physics teacher - whose name was Mr. Bader - called me down one day after physics class and said, "you look bored; I want to tell you something interesting. Then he told me something which I found absolutely fascinating, and have, since then, always found fascinating. Every time the subject comes up, I work on it. *Richard Feynman, in physics lectures, on the principle of least action.*

LEARNING IN THE FRAMEWORK OF PHYSICAL LAWS

Let us think of an agent which interacts with the learning environment with the purpose of optimizing its behavior with respect to a prescribed protocol on the exchange of information with the environment. In so doing, whenever we agree on the purpose of the agent and on the protocol, we rely on the principle of not to make any further assumption on the agent. We can think of learning processes as physical laws and we can try to capture their essence within the variational framework. In machine learning, the variational approach was advocated by Poggio and Girosi [4]; later on, related investigations gave rise to the theory of kernel machines. However, those developments are based on a very limited communication protocol, which only involves learning from a finite collection of examples. We advocate a new direction in which the agent is exposed to the interaction with the environment in a way that very much resembles what happens in child learning, where sub-symbolic and symbolic learning follows a critical sequential path. Like for physical laws, where variational principles ensure grace and leads to capture the simplicity and elegance of natural behavior, in cognitive science variational principles gives rise to kernel machines that provide an effective model of sub-symbolic tasks like those related to sensimotor and pre-operational stages in children. We can keep the same framework to impose a *semantic-based regularization* [5] to relate semantic attributes so as to capture higher levels of cognition. Unlike what arises from kernel machines, a remarkable feature of human learning is that, as the time goes by, we enter a stage of development in which the communication protocol starts involving rules in addition to examples of concepts. We can think of those rules as constraints on a number of semantic attributes of the learning task. Let $f : X \subset \mathbb{R}^d \to \mathbb{R}^p$ be the function associated with the learning agent which learns from a collection of examples and constraints. As usual, the examples are picked up from $\mathcal{E} \subset X \times \mathbb{R}^p$, while the constraints can be compactly expressed by $\Phi : X \times \mathbb{R}^p \to \mathbb{R}^q$, where $q < p$. The learner is supposed to fit the training set $\mathcal{E}$ and fulfill the constraints, that is to discover $f(x) = [f_1(x), \ldots, f_p(x)]'$ such that $\Phi(x, f(x)) \geq 0$ (or $\Phi(x, f(x)) = 0$). If we penalize examples and constraints in the same way by the classic hinge function $\ell_h(\cdot)$, we end up in the associated unconstrained variational problem

$$E(f) = \sum_{i=1}^{\ell} \sum_{j=1}^{p} \ell_h(x_i, y_{i,j}, f_j(x_i)) + \lambda \sum_{j=1}^{p} \int_X \parallel P f_j(x) \parallel^2 p(x) dx + \gamma \int_X \sum_{h=1}^{q} \ell_h(\phi_h(x, f(x))) p(x) dx,$$

where P any positive pseudo-differential operator[1] as discussed in [4]. This variational problem resembles the one which gives rise to kernel machines, but adding the constraint term make it significantly harder. This is especially true when the agent is asked to incorporate deep rules on semantic attributes, since they give rise to a hard non-linear variational problem for which the representer theorem, that makes it possible the reduction of the optimization to finite dimensions, does not hold any more. The collapse of dimensions (see e.g. [6], pp.144–146) is still possible provided that we approximate the penalty term as follows:

$$\gamma \int_X \sum_{h=1}^{q} \ell(\phi_h(x, f(x))) p(x) dx \approx \gamma \sum_{h=1}^{q} \sum_{\alpha=1}^{m} \ell(\phi_h(x_\alpha, f(x_\alpha))) \mu(x_\alpha) = \rho \sum_{h=1}^{q} \sum_{\alpha=1}^{m} \ell(\phi_h(x_\alpha, f(x_\alpha))),$$

[1] The green function of the psuedo-differential operators are simply the kernels used in kernel machines.

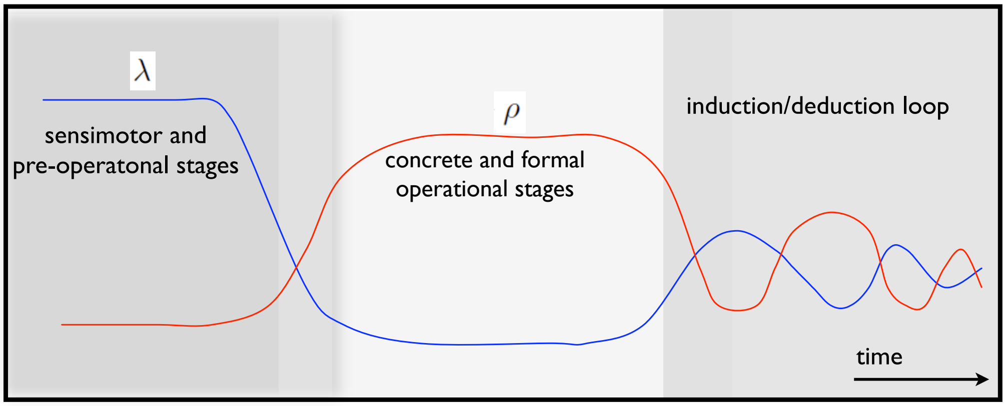

Figure 1: Insights from Piaget's cognitive stages to face the complexity of learning. The adoption of stages and the induction/deduction loop helps getting around local minima and sampling the penalty to enforce the constraints.

where $\mu(x_k)$ is the measure of the general k-th portion of a tessellation on X centered on m points, generally different from the training examples, and $\rho = \gamma\mu$, where μ is the weighted average of the measure $\mu(x_k)$. Unlike for kernel machines, however, once we plug the kernel expansion dictated by the representer theorem into the index $E(f)$, the resulting function is not convex and the corresponding optimization can be hard. In addition, the adopted approximation of the penalty term is based on a blind choice of m points, which might led to an unsatisfactory verification of the constraints. Piaget's cognitive developmental theory sheds light on the solution of the variational problem: We can start enforcing sub-symbolic learning by neglecting the constraint penalty term as shown in Fig. 1. To some extent, this corresponds with sensimotor an pre-operational stages in child development, where the focus is on perceptual tasks. Afterwards, the regularization parameters λ and ρ invert their trend so as to enforce the constraints, and the emphasis is shifted to semantic regularization. This late optimization of the constraint penalty term benefits from the approximation yielded during the first stage, so as the corresponding minimization begins from a point that is significantly closer to the overall optimum than trivial random start. Moreover, the approximate solution discovered at the end of the first stage is a precious source for any smart sampling of $f(x)$ to be used for the approximation of the constraint penalty term. This makes it possible to enforce rules so as to develop a behavior that partly reminds of concrete and formal operational stages in children. Interestingly, policies for getting around local minima and accurate techniques for the approximation of the constraint penalty term are likely to be discovered when considering that most interesting cognitive tasks are neither entirely acquired by induction neither by deduction, but they on their combined propagation (Fig. 1). This learning cycle in which λ and ρ exhibit an alternative behavior has nice cognitive roots in teaching strategies and student knowledge acquisition [7]. Not only are stages important, but also the sequence of the presented concepts [8, 9]. In kernel machines, the adoption of the hinge loss function $\ell_h(\cdot)$ leads to the striking simplification of restricting the kernel expansion to support vector. Likewise, semantic-based regularization leads to *support constraints*, but their meaning does not limit to the restriction of the involved constraints, since their structure suggests using appropriate ordering of presentation for on-line learning.

The analysis of the agent interaction with an environment that also offers relations on semantic attributes can be described by physical laws within the variational framework, thus extending classic Tikhnov regularization to semantic-based regularization. While this can shed some light on the comprehension of cognitive stages in general, the suggested variational approach does not address a number of issues in child learning including syntactic structure in the language, focus of attention, consciousness, and emotions that all play a key role in human cognition. However, in this paper we claim that the way biology breaks complexity, by dictating stages in the child development, seems to be an instance of general variational principles from which the need of development stages and of the induction/deduction loop clearly emerge as the outcome of simple interactions with examples and constraints that express relations on semantic attributes. Finally, the secret for breaking the complexity of learning for any agent that fruitfully interacts with a learning environment in which ”rules” circulates in addition to examples of concepts is to adopt a stage-based learning process in which semantic regularization, to incorporate constraints, takes place only after a first purely inductive stage based on classic Tikhnov regularization.

References

[1] Inhelder, B., Piaget, J.: The Growth of Logical Thinking from Childhood to Adolescence. Basic Books, New York (1958)

[2] Piaget, J.: La psychologie de l'intelligence. Armand Colin, Paris (1961)

[3] Gorse, D., Shepherd, A.J., Taylor, J.: The new era in supervised learning. Neural Networks **10** (1997) 343–352

[4] Poggio, T., Girosi, F.: Networks for approximation and learning. Proceedings of the IEEE **78** (1990) 1481–1497

[5] Gori, M.: Semantic-based regularizaton. Technical report, University of Siena (2009)

[6] Hastie, T., Tibshirani, R., Friedman, J.: The elements of statistical learning: Data mining, inference, and prediction. Springer (2001)

[7] Karplus, R.: Teaching for the development of reasoning. Research in Science Education **10** (1980) 1–9

[8] Cornuejols, A.: The necessity of order in machine learning: Is order in order? In: In order to learn: How the sequences of topics affect learning - J. Ritter Nerb, E. Lehtinen, and T. OShea (Eds.), Oxford University Press,. (2007) 41–56

[9] Ritter, F.E., Nerb, J., Lehtinen, E.: (Eds) In order to learn: How the sequence of topics influences learning. Oxford University Press (2007)