Cognitive meta-learning of syntactically inferred concepts

Article

Accepted version

S. Gaglio, G. Lo Re, M. Ortolani

In Frontiers in Artificial Intelligence and Applications, 2011, pp. 118-123

Cognitive Meta-learning of Syntactically Inferred Concepts

Salvatore GAGLIO [a,b], Giuseppe LO RE [a] and Marco ORTOLANI [a,1]
[a] *DICGIM, University of Palermo, Italy*
[b] *ICAR-CNR, Italian National Research Council*

Abstract. This paper outlines a proposal for a two-level cognitive architecture reproducing the process of abstract thinking in human beings. The key idea is the use of a level devoted to the extraction of compact representation for basic concepts, with additional syntactic inference carried on at a meta-level, in order to provide generalization. Higher-level concepts are inferred according to a principle of *simplicity*, consistent with Kolmogorov complexity, and merged back into the lower level in order to widen the underlying knowledge base.

Keywords. Cognitive learning, Grammar Inference, Kolmogorov complexity.

Introduction

One of the most distinctive traits of human intelligence is the ability of recognizing similarities across seemingly different contexts; loosely speaking, the core of human understanding involves matching observed events to categories, which may represent the generalization of specific individual instances onto comprehensive representative concepts. Such mechanisms favor the arising of surprisingly complex behavior in biological systems, which has been the subject for investigation aimed at devising reliable models for the human brain, as well as at discovering effective approaches to its automatization.

A striking example of the elaborate operational organization of the brain is represented, for instance, by the ability of perceiving complex visual scenes, where the sensory perception does not arise as a mere sum of elementary stimuli, but rather as a complex process of transformation of simpler pieces of information [1]. This is well modeled by the *Gestalt* theory of mind [2], which considers the human brain as a holistic, parallel machine with self-organizing capabilities; in particular, the essence of the cognitive abilities is the capacity of integrating current perceptive information into a coherent framework, and to merge new information with past experiences. In [3], for instance, the issue of measuring the complexity of the human brain is addressed by adopting a unified approach based on information theory; basically, the idea is to estimate the statistical mutual information exchanged between different neural areas in order to compute a complexity measure for neural activity, with high complexity characterizing systems that are both highly integrated and specialized.

This view naturally fits with the *connectionist* approach to artificially simulating and assessing the brain's functionalities, of which Artificial Neural Networks (ANNs) are the

[1]Corresponding Author: Marco Ortolani, DICGIM, University of Palermo, Viale delle Scienze, ed. 6 90128 Palermo, Italy; E-mail: marco.ortolani@unipa.it.

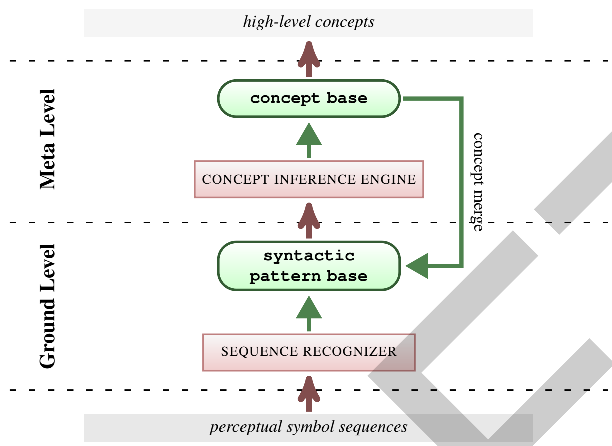

Figure 1. Block diagram of the overall system organization.

most notable example. The advantages of using such kind of methods are manifold, as they provide compact and computationally efficient representations, and appear to match closely the physiological structure of the brain; however, as noted by [4], a fundamental epistemological problem for ANNs is that even though an ANN system may have produced a satisfactory categorization, nonetheless it will likely fail to provide any natural explanation for that category, thus making it hard to attach a semantic value to it.

In the present work, we advocate the use of an historically alternative approach, namely the *symbolic* one, for reproducing some of the brain's basic functionalities; more specifically, we aim at modeling its abstraction capability to match sequences of perceptual inputs and recognize their underlying *syntactic* structure, which will be used to extract a more general pattern representing entire sequences. A remarkable advantage of using a syntactic approach is the self-explanatory quality of the obtained patterns, and their intrinsic *generative* nature. On the other hand such approach has often been regarded as too rigid, with respect to the dynamics of the process of understanding; for this reason, we propose to extend the basic pattern recognition structure with a meta-level where syntactically extracted patterns may be aggregated, and novel patterns may emerge thanks to a guided inference; such higher-level pattern representative may thus be considered the equivalent of *concepts* in the human brain. We refer to the *cognitivist* viewpoint, following the considerations expressed by Gärdenfors in [5], in whose opinion meaning needs to be *perceptually grounded*; in other words, unlike the realistic approach which claims that meaning arises from a mapping between the language and the external world, the mental structures applied in cognitive semantics represent on their own the meaning of the perceived symbol sequences.

In our approach, the issue of meaning is addressed in terms of operational semantics, since concepts are identified with computational entities, such as Finite State Automata (FSA). The internal representation arising from the analysis of perceptions will be structured by triggering the selection of more general, simpler concepts from the basic syntactic structure recognized at the lower level, and to this purpose we have devised a system, whose two main constituting parts are represented by the block diagram in Figure 1; in particular, the backward arrow on the right side of the picture shows how higher-level concepts are somewhat *internalized* by being fed back into the lower level.

We claim that the principle steering the selection of the most representative concepts ought to be *simplicity*, according to the well known Occam's razor principle, which in our case will be modeled by taking into account the Kolmogorov complexity [6] of the produced concepts. The aim of the paper is thus to outline the design of a framework for extracting high-level concepts from sensory perceptions represented by sequences of symbols, according to a syntactic pattern recognition process, driven by higher-level inference of novel, more representative concepts. The framework aims to provide explicit representation of the abstraction process occurring in human brain.

After providing a brief summary on the relevant scientific background, the remainder of the paper will present an outline of the proposed cognitive architecture.

1. Scientific Context

The brain's capacity of integrating current perceptive information into a coherent framework, and to merge new information with past experiences is the core of our cognitive abilities; studies on this topic have fostered the development of the field known as cognitive science, devoted to the formulation of a computational theory of mind.

According to Gärdenfors [4], three levels of representation of knowledge are typically identified in cognitive science: the associationist (or subsymbolic) level, the conceptual level, and the symbolic level; in fact, most recent literature in machine learning, has favored associationism, and more specifically, connectionism as opposed to the earlier attempts to investigate symbolic approaches.

Connectionist systems, such as ANNs, consist of large numbers of simple and highly interconnected units, which process information in parallel; according to connectionism, cognitive processes should not be represented by the manipulation of symbols, but rather by the dynamics of the activity patterns in ANNs.

Such dynamics may be interpreted in terms of the interplay between *functional segregation* (the possibility for different brain regions to be activated by specific cognitive tasks or by specific stimuli), and *functional integration* (the ability to rapidly and coherently consolidate diverse signals in order to drive adaptive behavior), as proposed by the authors of [3]. In order to mediate between the two opposing requirements, the same authors propose a unified approach based on information theory, aiming at estimating the statistical mutual information exchanged between different neural areas in order to compute a complexity measure for neural activity, with high complexity characterizing systems that are both highly integrated and specialized.

Connectionist approaches are characterized by the effort to model the intrinsic adaptiveness of the brain to the diversity of the external stimuli by way of highly dynamic internal representations. A different view on the same issue comes from the symbolic approach. According to the seminal article by Newell and Simon [7], "the central tenet of the symbolic paradigm is that representing and processing information essentially consists of symbol manipulation according to explicit rules." Even though this may sound too rigid, it carries the remarkable advantage of directly providing a compact representation as well as an *interpretation* of the input; moreover, symbolic analysis may be implemented efficiently since manipulations of symbols are performed regardless of the semantic content of the symbols. On the other hand, this has been considered a serious drawback of symbolic systems, also with respect to the issue of symbol grounding.

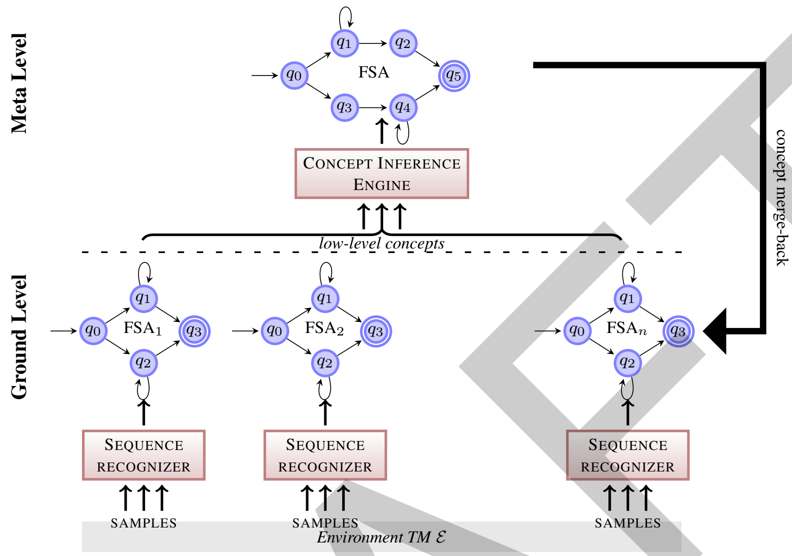

Figure 2. Scheme for the syntactic cognitive architecture.

In [8], the authors point out that biological complexity needs evolution to be modeled reliably, and in spite of its underlying rigidity, the symbolic approach is able to take into account the intrinsic need for adaptiveness. The internal representations may in this case be formulated in terms of automata, or Turing machines (TM), and in [9], a general formulation of an AI system is provided in terms of TMs, which allows to state a precise formalization of the principles that ought to drive the selection of the "fittest" within a set of such machine in order to fulfill a given task has been provided, according to a criterion of simplicity, similar to the renowned Occam's razor. Basically, an AI system may be described in terms of an iterative agent with some internal state, which at each step k acts with output y_k on some environment, perceives some input x_k from the environment and updates its internal state. Focusing on the deterministic case, if input and output are represented by strings, the system can be modeled by a Turing machine (TM) $\mathcal{S}$, governed by a deterministic policy which determines the reaction to a perception; if the environment is also computable, it might be analogously modeled by a TM $\mathcal{E}$. As will be shown in the following, this formulation together with a precise, formalized notion of complexity may be used to describe a cognitive system with capabilities resembling human abstract thinking.

2. Inferring Complex Concepts from Sequences of Symbols

Following the theoretical framework outlined by Hutter in [9], we assume here that the environment from where perceptions are drawn is computable, hence representable as a Turing machine (TM). We consider sequences of symbols as the sensory input of our system and we aim to recognize the implicit patterns present in such sequences and to internalize them both as compact syntactical descriptors, and as more general concepts.

Figure 2 shows a more detailed representation of the syntactic cognitive architecture. The lower part of the picture shows the *ground level*, where sequences of symbols are recognized as what we may regard as *pattern primitives*, following the scheme of [10].

Such low-level concept representatives might be hard-coded as primitive concepts in our system; for the purpose of this discussion, we will assume that the environment is representable as a TM in one of its simplest forms and for simplicity's sake, we will restrict our analysis to the case of regular grammars, so that in fact the representation for a low-level concept will be a finite state automaton (FSA). FSA offer the twofold advantage of providing a compact representation for an entire category of perceptions, as well as an efficient implementation, in terms of memory, and time requirements. More importantly, automata are self-explanatory, and much more powerful than, for instance, ANNs in that regard, as they are not mere representatives of a category of samples, but rather *generators*.

Simulating human abstract thinking cannot disregard a natural component of evolution, so our "syntactic pattern base" (i.e. the set of the basic FSA collected so far) cannot be static; on the contrary it needs to adapt to new samples potentially belonging to previously unseen sequences, as well as to the emergence of more general sequences, i.e. more general FSA representing higher-level concepts. In other words, the structure of the discovery process of new concepts tries to mimic the process of human thinking, in line with the cognitivist approach.

In our system, the choice of higher-level concepts will be driven by the *simplicity of description*; namely, as shown in the figure, we superimpose a *meta level* to the base syntactic inference level, and we consider a CONCEPT INFERENCE ENGINE whose purpose is to analyze low-level concepts as well as raw symbol sequences in order to attempt a further inference of "better", more general and more compact representations. At the meta level, syntactic learning is performed by carrying out grammatical inference on the lower level perceptions. Syntactic learning consists in the inference of the grammatical structure presumably underlying the provided samples, a process also known as *grammar induction*. Successful inference will produce a grammar providing a compact representation for class of samples which we basically deem belonging to the same conceptual category. Some effective methods for grammatical inference are reported in [10]. Basically, one can proceed by applying a simple inference *by enumeration* as initially proposed by [11], which relies on a partial ordering of a set of candidate grammars in order to choose the one better matching the samples. Alternatively, grammars may be inferred *by induction*, as suggested by Solomonoff [12]; in this case, the algorithm consists of a process of deletion/re-insertion of substrings from each string in the sequence in order to discover the recursive structure of the underlying grammar. Whatever the chosen method, Fu [10] proposes a general scheme for the grammatical inference process which relies on the use of the so called *Informant*, i.e. a complete information sequence drawn from the original language of the samples, to be used as a guidance for the inference process.

In order to fit the intuitive notion of simplicity of representation into our formalization (following the principle of Occam's razor) a useful concept is that of Kolmogorov complexity [6], also known as algorithmic entropy. The basic idea is that not only does a given symbol sequence need to be provided with a simple description, but the tool used to interpret that sequence (a TM, or an FSA in our case) also needs to be described in a simple way. We rely on Solomonoff Theorem [12], so that every possible description is to be formulated in terms of the Universal TM $\mathcal{U}$; with reference to [9] with slight simplifications, the notion of *complexity* of a concept, and consequently of the sequences it recognizes, might be expressed in terms of the Kolmogorov complexity as follows.

Let $\mathcal{U}$ be the reference universal Turing machine; the Kolmogorov complexity for a sequence x is defined as the shortest program p, for which $\mathcal{U}$ outputs x:

$$K(x) := \min_{p}\{length(p) : \mathcal{U}(p) = x\}.$$

Unfortunately, a straightforward use of $K(x)$ as a metric for complexity is not viable, since it is not a computable function; in our system we instead adopt a heuristic approach, and measure the complexity of a representation by the number of the states of the minimal DFA, equivalent to the considered FSA. Moreover, the CONCEPT INFERENCE ENGINE, at the meta level in Figure 2, is devoted to infer novel representations, based on previously extracted ones as well as on the original sequences; such new concepts will however be provided with a description formally similar to those coming from the lower level, and may thus be merged back into the ground level in order to be consolidated into the existing knowledge. Such procedure simulates the "hard coding" of newly acquired concepts into the lower level; the whole process may go on with an ever increasing structured and self-explanatory knowledge base.

3. Final Considerations

The paper presented an initial proposal for a cognitive architecture aiming at simulating human abstract thinking. Only the basic ideas have been sketched, and each part of the system needs to be analyzed in more detail, however the general description should convey the broad idea of representing the abstraction process thanks to a meta level where syntactic inference is performed.

A final note regards the flexibility of the proposed system; first of all, it is worth noting that we do not restrict long symbol sequences to trigger just one of the basic FSA, rather they will be recognized by more FSA at once, similarly to what happens to perceptions that may stimulate different concepts in the human brain. Finally, an additional degree of freedom for the system may be included by allowing for the presence of noisy samples; a first approach in this direction might be for instance the inclusion of a module for Bayesian inference of stochastic grammars in the CONCEPT INFERENCE ENGINE.

References

[1] Eric R. Kandel, James H. Schwartz, and Thomas M. Jessell, editors. *Principles of Neural Science*. McGraw-Hill, New York, fourth edition, 2000.

[2] K. Koffka. *Principles of Gestalt psychology*, volume 7. Psychology Press, 1999.

[3] G. Tononi, G.M. Edelman, and O. Sporns. Complexity and Coherency: Integrating Information in the Brain. *Trends in cognitive sciences*, 2(12):474–484, 1998.

[4] Peter Gärdenfors. *Conceptual Spaces: The Geometry of Thought*. The MIT Press, 2004.

[5] Peter Gärdenfors. Some Tenets of Cognitive Semantics. *Cognitive semantics: meaning and cognition*, pages 19–36, 1999.

[6] A. N. Kolmogorov. Three Approaches to the Quantitative Definition of Information. *Problems of information transmission*, 1(1):1–7, 1965.

[7] A. Newell and H.A. Simon. Computer Science as Empirical Inquiry: Symbols and Search. *Communications of the ACM*, 19(3):113–126, 1976.

[8] R. A. Wilson and F. C. Keil. *The MIT encyclopedia of the cognitive sciences*. The MIT Press, 2001.

[9] Marcus Hutter. *Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability*. Springer Verlag, 2005.

[10] K. S. Fu. *Syntactic Methods in Pattern Recognition*. Academic Press New York, 1974.

[11] J. A. Feldman. First Thoughts on Grammatical Inference. *Artificial Intelligence Memorandum*, 1967.

[12] R. J. Solomonoff. A formal Theory of Inductive Inference. Part I and II. *Information and control*, 7(1):1–22, 1964.