Syntactic learning via Topos Theory

Olivia Caramello

(University of Insubria - Como, and Grothendieck Institute)

New AI Theory workshop
RL China, 24 November 2023

The key features of true intelligence

In artificial intelligence and beyond, unification and computation are two facets of the same phenomenon: the more there is understanding, the less ‘brute force’ is needed and calculations can be simplified.

One should thus possibly design artificial learning systems in such a way that computations become as conceptually inspired and enlightening as possible, and, most importantly, meaningful, to the human mind. In particular, we should shift to a new conception of information which is based on semantics and invariants rather than on the classical set-theoretic foundations for mathematics.

One should also aim for strong forms of modularity and continuity ensuring, in particular, technical flexibility, adaptability and resilience capabilities of the resulting systems.

Despite the spectacular advances in deep learning systems of the last years, we are still far from a form of artificial intelligence enjoying these features, which in fact are quite characteristic of human intelligence.

Unification and computation

A general principle is that one should aim for the smallest possible number of general methods (*unification*) and for the greatest possible number of concrete ‘ingredients’ to which such methods can be applied (*computational part*).

This goes hand in hand with the development of conceptual architectures which should embody a small number of fundamental principles and be as technically flexible as possible in relation to the applications that they guide and orient.

This relies, in turn, on a continuous effort to try to isolate, in any situation, the conceptual part from the purely computational, ‘routinary’ or ‘mechanical’ one.

My aim in this talk is to suggest, on the basis of my mathematical experience, some ideas on how to orient research in order to try to achieve these goals. I will argue that Topos theory and Mathematical Logic are two very relevant subjects in connection with a development of new foundations for AI implementing the above-mentioned principles.

Mathematical Logic

- Mathematical logic is crucial since it is, in a sense, the "science of formalization". It notably allows one to study concepts and their relations in a fully rigorous way, to define the concept of mathematical theory and also, in a sense, to give a precise meaning to the intuitive notion of 'point of view'.
- Indeed, different ways of thinking about or of constructing a given object, or different knowledge representations for a given piece of information, translate into different formalizations, which can be studied in themselves as well as in relation with each other through mathematical methods.

  All of this ultimately relies on the fundamental distinction between syntax and semantics.
- Note that, with respect to natural languages, Logic has the advantage that its formal languages and theories can be investigated with powerful mathematical tools, notably including those of topos theory.

Topos theory

- Grothendieck toposes are abstract spaces on which the fundamental mathematical invariants are naturally defined.
- Topos theory thus provides a natural mathematical foundation for a theory of semantic information, where the latter is embodied by toposes and the articulation between a given meaning (or semantic content) and the different (syntactic) ways to express it corresponds to the technical duality between a topos and its different presentations.
- As it happens, since the times of my Ph.D. studies, I have developed a theory, namely the theory of 'toposes as bridges', providing a number of techniques allowing one to exploit precisely this duality for studying mathematical theories in a semantic and dynamical way, and for establishing 'bridges' across them.
- These 'bridges' have actually proved useful not only for connecting different theories with each other, but also for investigating a given theory from multiple points of view.

Classifying toposes

Grothendieck toposes are objects which are capable of capturing the essence of a great variety of different mathematical contexts. In particular, they can emboby the semantic content of a very wide class of theories:

Indeed, in the seventies, thanks to the work of a number of categorical logicians, notably including M. Makkai and G. Reyes, it was discovered that:

- With any mathematical theory $\mathbb{T}$ (of a very general form) one can canonically associate a topos $\mathscr{E}_{\mathbb{T}}$, called its classifying topos, which represents its 'semantical core'.
- Two given mathematical theories have the same classifying topos (up to equivalence) if and only if they have the same 'semantical core', that is, if and only if they are indistinguishable from a semantic viewpoint.
- Conversely, any topos is the classifying topos of some theory (in fact, of infinitely many theories).

Toposes for semantic information

- Toposes are thus objects which embody the semantic content of a wide class of theories, or the essence of a great number of formalized contexts.
- In particular, toposes can be meaningully associated with all the usual concepts used for knowledge representation (graphs, groupoids, posets, categories, topological spaces, simplicial complexes, fibrations, quantales, ...),
- It is therefore natural to develop a theory of semantic information based on toposes and on the invariants that one can define on them.
- The fundamental invariants of mathematical structures are actually invariants of toposes associated with these structures. Indeed, it is at the topos-theoretic level that invariants naturally live.
- The theory of “toposes as bridges” will thus provide a theory of relations between information (embodied by toposes) and the different knowledge representations for it (embodied by different presentations for toposes).

Empowering learning systems with logic

We plan to explore the possibility of empowering the functioning of artificial learning processes, in particular of deep neural networks, by building on the notion of mathematical proof:

- Note that any intelligent agent must, in order to get an effective understanding of (aspects of) the world, derive knowledge starting from certain 'sensory' inputs, which play a similar role to that of axioms for a mathematical theory, by following certain dynamical rules, which correspond to the inference rules of the logical system inside which the mathematical theory is formulated.
- As every mathematical theory can be enriched by the addition of new axioms, so the functioning of an agent can be updated by the integration of new information which becomes available to it.
- The functioning of a learning system would thus correspond to a sequence of a mathematical theories, each of which more refined (that is, with more axioms, or fewer models) than the previous ones.

Learning processes via proofs

- In principle, any learning sequence, as any process 'approximating truth', is infinite, but for practical purposes one is normally satisfied by the result of a learning process when the last theory in the sequence leaves a degree of ambiguity which is sufficiently low for the desired applications.
- All the theories in the sequence should extend (that is, be defined over) the basic theory of the agent formalizing its essential features. More generally, every theory in the sequence can be seen as theory defined *over* each of the preceding ones. This corresponds, through the classifying topos construction, to a sequence of relative toposes.
- Note that any constraints embedded in the logical formalism (or integrated at some step of the sequence of theories) will allow to significantly reduce the space of parameters that the agent has to explore and hence correspondingly decrease the computational complexity of the learning process.

Syntactic learning

In order to practically implement the above ideas, one should, first of all, empower any learning system with large formal vocabularies that will serve for expressing the concepts that the system will learn from data.

The idea is to enforce the learning to take place at the abstract level of syntax, rather than at that of a particular semantics, as it currently happens.

While the vocabulary should be given at the outset, we could let the system discover any (non-already embedded) logical rules expressible in that vocabulary by itself, by using the usual techniques. Also, we should let it suggest enrichments of the vocabulary on the basis of invariances empirically discovered in the data, thereby achieving a form of 'emergence of concepts'.

In this way, we could obtain systems capable of inferring all sorts of 'syntactic rules' from data (e.g. the grammar rules of a language from a great amount of samples of texts, or the rules of a game from a big collection of matches, etc.): these rules could, of course, be of different nature and complexity (e.g., propositional, first-order, higher-order, etc.).

Making AI systems 'speak'

All of this goes in the direction of constructing logical languages for AI agents, allowing them, in a sense, to 'speak'. After all, how can we expect a system to learn in a robust way if we do not give it the possibility of reasoning linguistically? Think about how our learning, as human beings, would be impaired without the possibility of expressing, testing and communicating our ideas by using languages.

Think also about the process of learning of a natural language. Little babies have no other possibility to learn a language than to merely rely on data (note, however, that their brain has already a lot of *a priori structures* which are used to organize and categorize the knowledge that they gather from the environment). Still, their knowledge of the language remains fragile until they are brought into contact, notably at school, with grammar, which represents syntax in this context.

The role of grammar is crucial in making *explicit* a lot of the *implicit* which had been accumulated in the previous 'bottop-up' learning process, and in bringing the understanding to a higher level, notably in terms of explainability.

Uncovering hidden structures and rules

Syntax represents the ‘skeleton’ of structures, whose concrete manifestations we access through data. Trying to ‘lift’ from a particular concrete setting to the level of syntax represents an abstract form of understanding which enjoys much more relisience and adaptability than the usual forms of statistically-based learning.

The deep reasons for the regularities that we may observe in concrete contexts actually live at the syntactic level (think, for instance, of motives in Algebraic Geometry, or to definability and preservation theorems in Logic).

Philosophically speaking, we need to teach learning systems to lift from the phenomenological level (of concrete manifestations of topos invariants) to the ontological one (of toposes themselves, regarded as classifying toposes of logical theories).

In conclusion, we advocate for an integration of a ‘bottop-up’ approach to artificial learning, such as the one which is dominant today, with a ‘top-down’ one based on logic and topos invariants.

For further reading

O. Caramello
Grothendieck toposes as unifying 'bridges': a mathematical morphogenesis,
in the Springer book *Philosophy of Mathematics. Objects, Structures, and Logics*.

O. Caramello
Grothendieck toposes as unifying 'bridges' in Mathematics,
Mémoire d'habilitation à diriger des recherches,
Université de Paris 7, 2016,
available from my website www.oliviacaramello.com.

O. Caramello
Theories, Sites, Toposes: Relating and studying mathematical theories through topos-theoretic 'bridges',
Oxford University Press, 2017.