



The new frontiers of AI in the arena of behavioral economics

Mario Rasetti^{1,2}

Received: 31 December 2019 / Accepted: 31 December 2019 / Published online: 24 April 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

There is a revolution going on, the digital revolution, engulfing all of us. Today the quantity of data that humans produce is huge: according to the TLC–Data Flow 2019 (Cisco Statistics Report), 5 zb in 2019 (1 zb = 10^{21} bytes; a thousand billion gigabytes!), that will become 15 zb in 2022. And it doubles now every year: in 2019 we generated as much data as had been produced in the entire history of humanity up to 2018. With the IoT catching up, in $2 \div 3$ years 150 billion sensors will be connected in a huge network, among themselves and with the humans: then the data doubling time will be 12 h.

The only way to deal with such a Tsunami and to extract value from these massive data sets is artificial intelligence (AI), which can grasp and organize the enormous amount of correlations hidden in data.

In the arena of behavioral economics the keystone is behavior. In economics it bears on all sorts of proxies: social networks (opinions, preferences, beliefs); personal digital census (education, ethnicity, age group, sexual orientation); lifestyle; health; use of e-commerce channels.

In an ideal definition of consumer's behavior, failures to repay debts, framing perspectives or price anchors should not have any bearing on choices, and decisions would be merely the result of a careful weighing of costs and benefits, informed exclusively by concrete, well-defined needs and preferences, with every decision rational. Herbert Simon's concept of 'bounded rationality' dismantles such definition, bringing into play the notion that consumers' minds (behavior) must be understood relative to the environment in which they evolved; thus decisions are not always optimal, even because human information processing has other severe restrictions, due both to incomplete information and knowledge and limits to computational and logical capacities. Behavioral Economics Theory assumes that people (consumers) are boundedly rational agents, with a limited ability to process information.

✉ Mario Rasetti
mario.rasetti@isi.it

¹ ISI Foundation, Torino, Italy

² ISI Global Science Foundation, New York, NY, USA

Exploring just how available information affects the quality and outcome of decisions, and what happens in situations where people avoid information altogether, Richard Thaler coined the concept of mental accounting: people think of value in relative rather than absolute terms. They derive their pleasure not just from an object's value, but from the quality of the deal as well: its transaction utility. Consumers tend to work with the totality of their mental accounts: personal experience, reliable information, prompt feedback are the key factors that enable them to make good decisions; yet information avoidance takes place, circumstances in which people choose not to secure knowledge, even if freely available. Deliberate information avoidance has many aspects: physical disregard, inattention, biased interpretation of information; even conscious oblivion.

Nowadays, however, more and more much of the decisions are data-based; made either by humans with the assistance of machine intelligence or wholly by AI machines. It is reasonable to assume that AI may reduce the impact of bounded rationality, as AI processes reduce information asymmetry in the market and improve decision-making, thus making markets more rational. The open question is whether the use of AI in the market in applications such as online trading and decision-making may change economic theories, having on them an impact bearing on issues such as rational choice and expectations, computational thinking, portfolio optimization, counterfactual reasoning.

In what follows two novel tools of AI will be outlined, which should be able to cast a new light on all these questions: Topological data analysis (TDA) and constructor theory based machine learning (CT-ML); pointing out as well to the intrinsic limits that the AI approach could possibly exhibit (decidability of learnability). What we require AI to be able to do in the context of behavioral economics is: behavior analysis and forecast (cognitive science-based); support to decision-making (intelligence of processes, optimization of choices, predictive strategies, cognitive analytics); processing of 'languages' (natural as well as artificial); optimization policies.

With TDA's algebraic topology methods, integrated within the idea of treating data sets as spaces (their key property is that these are not vector spaces, but topological spaces) whose 'shape' is relevant, have progressively earned pivotal interest in data analytics. The reason of their success is that topological measures and observables are by construction very robust, and that moreover they permit to capture explicitly interactions between more than pairs of agents (nodes), thus providing a framework to describe, quantify, compare the global shape of arbitrary data spaces. This is crucial because virtually all interesting complex systems can be thought of as living in either configuration or phase spaces, including those that can be approximately described using finite datasets in terms of simplices. The two main concepts used to achieve this are 'persistent homology' and 'topological simplification'.

Persistent homology encodes the shape of topological spaces by progressively finer approximations—higher order analogs of links between nodes—in a network able to describe explicitly interactions of more than two agents at the same time. It allows us to identify (and reduce) noise vs. signal. The process emphasizes those topological features in increasing dimensions (one-dimensional cycles, three dimensional cavities, etc.), that survive through the sequence and therefore characterize

the shape of the dataset, letting us compare in a principled way arbitrary spaces with different dimensions, number of points, shape (invariants), etc. We can thus study the shape of correlation spaces among data space regions and how such shape changes. Functional, global, and localized homological information can all be used to track the system evolution in time and fingerprint individual subjects.

Topological simplification (known as Mapper, from its most famous algorithm) is a topological dimensionality reduction scheme, aimed to extracting low-dimensional simplicial-complex backbones from high-dimensional datasets, as it is possible to use this topological information to build a topological skeleton able to highlight dissimilarities both in structure and function of different behavioral pathways. This can be further leveraged to build a ‘topologically informed’ map of feature spaces, thus improving and stream-lining the selection of features important for classifications in such spaces (e.g., equivalence classes of correlation patterns).

On the other hand topological descriptions are equally useful in understanding artificial neural networks (ANN) and their capacity to learn new tasks. Topological methods have been realized to allow NN to take advantage of homological descriptors to better detect or craft adversarial attacks by exploiting the topology of learned manifolds, and to improve the interpretability of what actually happens inside NN as they learn to perform complex tasks. The crossover between topology, neuroscience and artificial intelligence occurs as the capacities of neural networks, like those of the human connectome, reside in how they represent data spaces internally, just like brain functions are encoded in functional patterns—a well defined problem of comparison of spaces. Topological invariants provide thus a common thread and a robust tool to understand both cognitive and behavioral processes and AI.

Constructor theory (CT) is a visionary extension of John von Neumann’s notion of ‘universal constructor’, a self-replicating machine in a cellular automata environment designed in the ‘40s, without a computer (the details were published in von Neumann’s book *Theory of Self-Reproducing Automata*, in 1966—completed by Arthur W. Burks after von Neumann’s death). Revived and fully (and rigorously) reformulated by David Deutsch and Chiara Marletto, CT was recently used to construct Information Theory (IT) completely and solely in terms of which transformations of the ground physical systems may occur and which may not (which is in a nutshell what constructor theory does: the basic principle of constructor theory is indeed that “All subsidiary theories are expressible entirely in terms of ‘statements’ about which physical transformations are «possible» and which are «impossible», and why.”). CT regards science—even IT—not merely as an enterprise for the purpose of making predictions, but as an enterprise for discovering what the world is really like, how it behaves and why.

A notorious problem with defining information within physics is that information is thought of as fully abstract: the theory of computation as developed by Alan Turing regarded computers and the information they manipulate in purely abstract terms as mathematical objects. One must realize instead that information is physical and that there is no such thing as an abstract computer: only a physical object can compute. Though it may include laws of physics that are only conjectured, CT-IT does not regard information as an a priori mathematical or logical concept, but as something whose nature and properties are determined by the laws of physics alone.

For this it does not suffer from the circularity at the roots of existing IT, namely that information and distinguishability are each defined in terms of the other. Thus CT reveals itself as the natural tool to proceed toward a true ‘ML Theory’.

Machine learning is the branch of AI concerning the construction and study of systems that can learn from data. Its core is the capacity of representing data instances and functions evaluated on these instances in such a way as to allow for recognition and construction of the method the system will perform with on different data instances. Keynote is the algorithm’s ability to perform accurately on new, previously unseen examples after having trained on a learning data set. In other words, the core goal of a learner machine is to generalize from its experience. The training results are probability distributions obtained from a reduced scale experience on the data set, while the learner’s task is to extract something more general, so as to produce useful predictions in new cases. One can say that ML focuses on the discovery of previously unknown global properties of the dataset.

As for IT, the processes taking place when ML operates can be pretty well represented in the frame of CT, as they do share a common physical frame: neural networks, be they natural or artificial. ANNs, brick circuits of AI machines, aim to mimic the human brain based on the concept that one way to think about the rational brain is that it works by accreting smaller abstractions into larger ones. Complexity of ‘thought’, in this view, is measured by the range of smaller abstractions one can draw on, as well as by the number of times one can combine lower-level abstractions into higher-level abstractions.

On the other hand, as more and more AI mediates our social, cultural, economic, political interactions, understanding the behavior of AI systems is crucial to our ability to control their actions, crop their benefits, minimize the harm they can do. This is what makes of AI the natural tool to deal with behavioral economics, even though a stronger scientific research agenda focus on machine behavior and interactive computing is badly necessary.

Giving a sound mathematical foundations to ML through CT (the technical tool is category theory) will progressively improve our understanding and provide us with novel principles and frameworks to design new learning paradigms; in particular ‘no go’ theorems. This bears on the fact that also ML cannot escape the curse that in any advantages of mathematics have a cost. In 1931, Kurt Gödel showed that in any system of axioms expressive enough to model arithmetic, for some true statements their truth is unprovable. Successively, it was shown that the Continuum Hypothesis (CH)—which states that no set of distinct objects has a size larger than that of the integers but smaller than that of the real numbers—cannot be proved nor refuted using the standard axioms of mathematics.

ML does not escape the fate of Gödel’s incompleteness theorems. Recently Shai Ben-David et al., resorting to the equivalence between learnability and compression, which implies that the solution to the respective optimization problem may be isomorphic to the proof of CH, succeeded in constructing scenarios proving that learnability may be undecidable in the sense of Gödel. Of course, identifying the learnable is (it must be) a fundamental goal of ML: but to achieve it, one needs a robust mathematical framework, supporting the formal treatment of learnability. Conventional paradigms of ML fail to do this, as learnability cannot always be decided by

standard axioms of mathematics, which are unable to provide any dimension-like quantity characterizing learnability in full generality. We argue that redefining such paradigms within the boundaries, rules and constraints of CT and TDA may lead to define efficiently such quantity.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.