Pattern Recognition Letters 000 (2015) 1–7

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Semantics of object representation in machine learning ☆

Birkan Tunç*

Center for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, PA 19104, USA

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Philosophy of machine learning
Object representation
Idealization
Abstraction
Latent variables

ABSTRACT

With the growing success of machine learning, both researchers and philosophers have recently regained their interest in the foundational problems of statistical learning. The cooperation between philosophy and machine learning has been recognized to be mutually beneficial that may provide fundamental shift in the paradigms of both camps. In this paper, a unidirectional interaction between philosophy and machine learning is considered. This type of interaction becomes necessary as we reflect upon the practical implications of the model construction. To this extent, I review a specific set of contributions of philosophy to machine learning in determining epistemic standing of object representation and algorithmic design. I discuss three aspects of statistical models, pertaining to semantics of object representation, namely idealization (simplifying properties of an object), abstraction (representing an object with another object that is easier to handle), and use of latent variables. I argue to what extent these aspects necessitate philosophical attention to justify their practical use. To this end, I elucidate different philosophical concepts that are utilized by researchers mostly tacitly when dealing with uncertainties in features and their functional relationships. This is expected to help pave the way for further investigations on semantics of object representation in machine learning.

1. Introduction

Child of a frog is a frog: as an inevitable consequence of the overall disconnection between science and philosophy, machine learning[1] researchers have eventually found it easier and more efficient to dismiss most of the foundational considerations of the domain. It is mainly to meet the immediate needs of practical applications, by concentrating only on the algorithmic and the technical aspects. Machine learning has therefore become associated with a peculiar vocabulary and progressed to develop a set of methodologies mostly without any direct correspondence to philosophical concepts.

Recently, however, with the growing success of machine learning, both researchers and philosophers have regained their interest in the foundational problems of statistical learning. The co-operation between these camps has been recognized to be mutually beneficial [4,55,56]. According to the proponents of such an interaction, the statistical/computational learning paradigms may provide new insights into the sources of knowledge and even epistemological aspects of the scientific methodology, whilst machine learning itself may favor new conceptual developments. Establishing a synthesis of the methodological authority of these two domains may provide stronger foundations that future machine learning studies will rely upon. Moreover, a dynamic and mutual interaction between domains may provide fundamental shifts in the paradigms of both camps.

The main goal of this paper is to elucidate how objects of real world are represented in machine learning studies and to show how philosophical consideration, pertaining to the semantics of object representation, can contribute with determining epistemic standing of devices that are used commonly by researchers. I review three such devices utilized in data modeling, namely *idealization* (simplifying properties of an object), *abstraction* (representing an object with another object that is easier to handle), and *latent variables*. I argue that these tools necessitate philosophical attention to justify their use by researchers. The scope of the paper thereby is limited to the discussions on the epistemic value of findings such as putative functional forms of interactions between observations. Below, I review other possible communications between philosophy and machine learning, in order to put the contributions of this paper in perspective.

Statistical learning can be defined as inferring unknown functional relationships between input data points and their continuous labels (*e.g.* regression), their discrete labels (*e.g.* classification) or their densities (*e.g.* clustering) [51]. The final product of a learning algorithm is the estimated function that can be used to predict the labels or densities of new observations. Interactions between philosophy and statistical learning, in its early days, began with the philosophical enquires on artificial intelligence and statistical data analysis [22,49,50]. Although foundations of statistics and inductive learning had been exhaustively studied since David Hume's "An Enquiry

☆ This paper has been recommended for acceptance by G. Sanniti di Baja.

* Tel.: +1 215 349 8694.
E-mail address: Birkan.Tunc@uphs.upenn.edu

[1] From the perspective I embrace in this article, *machine learning*, *statistical learning*, and *pattern recognition* can be (and are) used interchangeably.

http://dx.doi.org/10.1016/j.patrec.2015.03.016
0167-8655/

Concerning Human Understanding", [26] by various authors [23,24,47], machine learning became an uncontroversial subject matter of philosophy with the emergence of the artificial neural networks [44,48] as tools of cognitive studies, following a movement called *connectionism* [20].

In its contemporary form, the philosophical problems in machine learning are still mostly concerned with the foundational problems such as problem of induction [23]. The deliberations at the level of foundations mostly deal with the epistemic value of inductive inference. Rival implications of the frequentist and the Bayesian approaches and their potential problems become apparent at this level. In the literature, the interaction between philosophy and statistical learning has been predominantly studied (maybe not explicitly) at this level [15,24]. Besides problem of induction, other foundational issues such as essentialism *vs.* non-essentialism distinction have attracted some attention in machine learning community, in the context of similarity based learning [39].

Recently, another type of interaction between domains has been also advocated. Following the broader tendency to *naturalize* philosophy [8,21,41], statistical learning has been suggested as a means of improving several aspects of philosophy of science [13,30]. For example, the Vapnik−Chervonenkis (VC) dimension [51] has been offered as a useful extension for the Popperian concept of falsifiability [13]. As a rather extreme example, statistical learning has been also suggested to be considered as a new way of doing philosophy of science [30].

In this paper, only a unidirectional interaction between philosophy and machine learning is considered, without refuting other possible connections. This type of interaction becomes necessary insofar as we reflect upon the practical implications of the model construction. I advocate the idea that machine learning needs to be cultivated with the vocabulary of philosophy to extend the range of questions that are raised when evaluating various aspects of machine learning, pertaining to data representation. This will eventually enable us to study correspondence between nature and the structure of real entities that are modeled and the properties of corresponding mathematical objects employed in machine learning algorithms. Acquiring a new lexicon from philosophy of science and epistemology such as idealization, abstraction, and theoretical terms will help pave the way for further investigations on semantics of algorithmic design and their justification.

In subsequent sections, I will first try to justify the need for philosophical attention directed toward data modeling in machine learning. Then, after introducing several essential concepts from philosophy of science, I will discuss three aspects of machine learning studies that require us to refer to these philosophical concepts.

2. Why should philosophical aspects of data modeling be studied?

One of the main debates in epistemology and ethics is about the criteria for a "good" belief or action, which ultimately converges to a discussion on the notion of justification. In the context of epistemology, justification can be defined as the reasons to hold a belief. In the setting of this paper, justification may refer to the reasons to use any model design and object representation in our statistical analyses. From a broader perspective, the reasons to use a machine learning tool in practice were studied by Corfield [12]. He listed four types of justification that are offered by machine learning theorists for their inductive algorithms [12]. Among these four, the first two assess the absolute performance of an algorithm as quantified by the unbiased estimate and probabilistic bounds of the generalization error. Other two are related to the relative performance of an algorithm compared to the alternative algorithms and other ways the same algorithm might function. He concludes highlighting the need for philosophical attention to the types of justification used in machine learning, a subject that for the most part remains largely unexplored.

It is clear how such varieties of justification can help us to assess the expected accuracy of a proposed model, in the absolute or relative sense. On the other hand, they cannot be utilized to justify our reasoning behind the choices that determine the design of the model itself. In other words, although they are useful when an algorithm is treated as a *blackbox*, without any deliberation on the inner structure, another question still remains: does the model design reflect the structure of the studied system adequately? This is indeed a valid question, since today statistical learning is used ubiquitously in natural sciences to infer relations among variables of the studied systems and to make causal interpretations based on stipulated functional relationships.

Today, our advanced machine learning tools let us discover previously unknown interactions between observed/latent variables in almost any domain of inquiry. For example, we can easily define boundaries that separate different classes such as mental disorders. From a data analyst's perspective, it can be enough to show that the label of a test subject can be successfully assigned. Nevertheless, when it comes to understand the nature of the boundary that separates disorder from normality or the boundary that distinguishes one disorder from another, details of the model design should also be justified [45]. That is to say, insofar as we want to go beyond the utilitarian perspective of a "good" algorithm (*i.e.* assessing the expected accuracy of a proposed model), we must broaden need for the philosophical attention to the topics beyond the accuracy, consistency, and the relative performance compared to competitors. In this context, I believe that there are lots of questions to be answered regarding the propriety of data representations and their semantic implications.

Let us take representation of data points as an example. By our well-founded mathematical theories, we know exactly how to represent data of different types as points in vector spaces. Complex structures such as graphs or unstructured data such as text documents can be implicitly represented as elements of a reproducing kernel Hilbert space, an approach known as *kernel trick* [42]. Undoubtedly, the kernel trick enables us employ the thoroughly perfected instruments of linear algebra, thereby boosting the performance of any application, regardless of the nonlinearity of original features. However, without a proper understanding of the concept of abstraction, we cannot study the extent of the correspondence between real entities of the system and the vectors in this veiled Hilbert space, which is known to us only through inner products.

New problems arise as we start to challenge explanatory capacity of proposed learning tools. According to the causal account of explanation [46], a good explanation can be defined as one that exposes the conditions under which and in virtue of what the explanandum obtains. For example, to answer if a classification framework that is highly accurate in detecting cancer cells or segmenting tumor regions can explain the underlying pathophysiology of the disease, we need to show that it reveals possibly causal relationships between selected/extracted input features and the output label. But, can we rely on the relationships among variables that are mere mathematical abstractions or controversial latent variables while real goal is to reveal physical properties of cells? Even if we evade the problem related to physical adequacy of variables, the propriety of several other components such as data geometry, likelihoods, priors, and their interactions is still susceptible to the same sets of questions.

The interpretation of explanatory accounts such as causal nets becomes more complicated as we introduce latent variables, a concept that begs for further philosophical attention. An analogous example can be borrowed from the philosophy of science to illuminate the subject. In quantum mechanics, the practical efficacy of the wave function is now beyond any doubt. Theories built around the wave function are able to provide very precise and comprehensible interpretations of a large body of phenomena. On the other hand, the ontological adequacy of the wave function is still a controversial topic [35] *i.e.*