Journal of Applied Logic 13 (2015) 78–90

Contents lists available at ScienceDirect

Journal of Applied Logic

www.elsevier.com/locate/jal

Abduction: A categorical characterization

Fernando Tohmé [a], Gianluca Caterina [b], Rocco Gangle [b]

[a] INMABB, Conicet and UNS, 12 de Octubre y San Juan, Bahía Blanca, Argentina
[b] Center for Diagrammatic and Computational Philosophy, Endicott College, 376 Hale Street, Beverly, 01915, MA, USA

ARTICLE INFO

Article history:
Received 10 September 2014
Accepted 21 December 2014
Available online 7 January 2015

Keywords:
Abduction
Category-theoretical representation
Adjunction

ABSTRACT

Scientific knowledge is gained by the informed (on the basis of *theoretic* ideas and criteria) examination of data. This can be easily seen in the context of quantitative data, handled with statistical methods. Here we are interested in other forms of data analysis, although with the same goal of extracting meaningful information. The idea is that data should guide the construction of suitable models, which later may lead to the development of new theories. This kind of inference is called *abduction* and constitutes a central procedure called *Peircean qualitative induction*. In this paper we will present a category-theoretic representation of abduction based on the notion of *adjunction*, which highlights the fundamental fact that an abduction is the most efficient way of capturing the information obtained from a large body of evidence.

1. Introduction

A possible classification of scientific activities focuses on the kinds of inference mechanisms applied to gain further knowledge. Pure theoretical branches usually use mathematics and therefore *deductive* inference[1] while more empirically oriented ones use statistical (i.e. *inductive*) inference disguised in various forms. The question to be raised by a logician in the Peircean tradition is how to accommodate the third type of inference, *qualitative induction* or *abduction*.[2] To answer this question, let us note first that the key procedure in this kind of induction (which should be clearly distinguished from statistical induction) is the inference from *evidence* to *explanation*. That is, abduction does not predict which evidence should be observed given a theory—a deductive inference—nor does it build a general description (*prototype*) of the evidence—a statistical inference. Rather, in scientific matters, abduction is the reasoning process that helps

E-mail addresses: ftohme@criba.edu.ar (F. Tohmé), gcaterin@endicott.edu (G. Caterina), rgangle@endicott.edu (R. Gangle).

[1] On the other hand, the burden of mathematical activity lies in conjecturing possible results or arguments to prove statements. This is also an abductive reasoning, of a kind that will not be discussed here.

[2] For a full characterization of this type of induction see [30].

http://dx.doi.org/10.1016/j.jal.2014.12.004
1570-8683/

to find theoretical constructs providing plausible explanations for how the data drawn from the real world were in fact generated.

In the rest of this paper we will discuss the role of abductive reasoning and extend it to a more general line of inquiry: to find qualitative descriptions of the information found in crude data. Section 2 is devoted to presenting a discussion on the meaning of abduction. Section 3 presents a category-theoretical environment for abduction, seen as a functor among "data" and "theoretical" categories. Sections 4 and 5 characterize in two different, albeit related, ways the abduction functor, one in terms of an *a priori* selection of potential "theoretical" outcomes and the other on constraints on the admissible selections *given* the "data" inputs. Finally, Section 6 discusses the conclusions of this work and presents possible lines for further inquiry.

2. Abductive reasoning and data analysis

Peirce emphasized the importance of Kepler's example for understanding how abduction works. The German astronomer, working with the huge database of planetary observations collected by Tycho Brahe, used his knowledge of geometry to conjecture that the planets follow elliptic paths around the Sun [26]. Without this insightful result, Newtonian physics would not have been possible. Similarly, a great deal of scientific theorizing arises from the insights provided by the examination of data. Any scientist, faced with some data, always tries to detect a pattern. In our terms, she tries to perform an informal abductive inference.

One of the relevant contexts in which abduction could be applied is analogous to Kepler's example. Although it sounds rather obvious, let us emphasize that there is a gap between the formulation of a question to be answered through measurement and the *actual* measurement providing the right answer. This difference arises from the fact that problems are usually stated in qualitative terms while data can be quantitative. In consequence, rough data (which certainly includes the quantitative counterparts of qualitative concepts) must be organized according to the qualitative structure to be tested. That is, a correspondence between theory and data must be sought. So, for example, in many socially oriented disciplines there exists a crucial distinction between ordinal and cardinal magnitudes in the characterization of *preferences*. But once measurements are involved it is clear that the theoretical relational structure must be assumed to be homomorphic to a numerical structure [19].

This implies that if there exists a database of numerical observations about the behavior of a phenomenon or a system, we might want to infer the properties of the qualitative relational structure to which the given numerical structure is homeomorphic. Of course, this is impaired by many factors:

- The representation of the qualitative structure may not have a unambiguous syntactic characterization [6].
- Heterogeneity in the representation hampers the unification of data sources.[3]
- Even if the observations fall in a numerical scale, the real world is too noisy to ensure a neat description of phenomena under consideration.
- There are complexity issues that make it highly convenient to just look for approximations, instead of a characterization that may make sense of every detail.

These factors, which usually preclude a clear cut characterization of the observations, leave ample room for arbitrary differences. In this sense, the intuition and experience of the analyst determine the limits of arbitrariness. Yet the reasoning process that justifies the decisions actually made is not often made clear. More generally, since empirical scientists spend a great deal of their time looking for relations hidden in the data, the process they apply to uncover those relations cries out for clarification.

[3] On the heterogeneity of data see [38]. See also the discussions on the convenience of having *iconic* representations in [8].