International Journal of Approximate Reasoning 102 (2018) 1–20

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Explanatory relations in arbitrary logics based on satisfaction systems, cutting and retraction ☆

Marc Aiguier [a], Jamal Atif [b], Isabelle Bloch [c,*], Ramón Pino Pérez [d]

[a] MICS, Centrale Supelec, Université Paris-Saclay, France
[b] Université Paris-Dauphine, PSL Research University, CNRS, UMR 7243, LAMSADE, 75016 Paris, France
[c] LTCI, Télécom ParisTech, Université Paris-Saclay, Paris, France
[d] Departemento de Matemáticas, Facultad de Ciencias, Universidad de Los Andes, Mérida 5101, Venezuela

A R T I C L E   I N F O

Article history:
Received 28 February 2018
Received in revised form 6 June 2018
Accepted 31 July 2018
Available online 6 August 2018

Keywords:
Explanatory relation
Retraction
Cutting
Satisfaction systems

A B S T R A C T

The aim of this paper is to introduce a new framework for defining abductive reasoning operators based on a notion of retraction in arbitrary logics defined as satisfaction systems. We show how this framework leads to the design of explanatory relations satisfying properties of abductive reasoning, and discuss its application to several logics. This extends previous work on propositional logics where retraction was defined as a morphological erosion. Here weaker properties are required for retraction, leading to a larger set of suitable operators for abduction for different logics.

1. Introduction

Since its introduction by Charles Peirce in [34], abduction has motivated a large body of research in several scientific fields, e.g. philosophy of science, logics, law, artificial intelligence, to mention a few. Abduction, whatever the adopted view on its treatment, involves a background theory (T), an observation also called explanandum (φ), and an explanation (ψ). The observation may be seen as a surprising phenomenon that is inconsistent with the background theory. It may also be consistent with the background theory but not directly entailed by this theory, which is the case considered in this paper. Several constraints can be imposed on the explanations and on the process of their production. One can allow changing the background theory, or not, consider as non relevant explanations those that entail the observation on their own without engaging the background knowledge. Hence, several forms of abduction can be defined depending on the chosen criteria. Despite their divergence, most of these models agree to define abduction as an explanatory reasoning allowing us to infer the best explanations of an observation. This contributes to the field of explainable artificial intelligence. Explanatory relations, trying to model common sense and everyday reasoning, find applications in many domains, such as diagnosis [16,22], forensics [30], argumentation [11,12], language understanding [31], image understanding [4,41], etc. (it is out of the scope of this paper to describe applications exhaustively). Then, as a form of inference, several rationality postulates have been studied, that are more appropriate to govern the process of selecting the best explanations, e.g. [24,

☆ This paper is part of the Virtual special issue on Defeasible and Ampliative Reasoning, Edited by Ivan Varzinczak, Richard Booth, Giovanni Casini, Szymon Klarman and Gilles Richard.

* Corresponding author.

E-mail addresses: marc.aiguier@centralesupelec.fr (M. Aiguier), jamal.atif@dauphine.fr (J. Atif), isabelle.bloch@telecom-paristech.fr (I. Bloch), pino@ula.ve (R. Pino Pérez).

https://doi.org/10.1016/j.ijar.2018.07.014
0888-613X/

35]. From a computational point of view, a very large number of papers has tackled the definition of abductive procedures, mainly in propositional logics. An attractive approach, governed by what is called the AKM model, is based on semantic tableaux tailored for particular logics (e.g. propositional logics [3], first order and modal logics [14,15]), which was the basis for several extensions (e.g. [5,22,28,37]). In these approaches, the explanatory reasoning process is split into two stages: (i) generating a set of hypotheses from the formulas that allow closing the open branches in the tableau constructed from $(T \cup \{\neg\varphi\})$, and (ii) selecting the preferred solutions from this plain set by considering some of the criteria mentioned above.

Our aim in this paper is to introduce a new framework for defining abductive reasoning operators in arbitrary logics in the framework of satisfaction systems. To this end, we propose a new notion of cutting, from which operators of retraction are derived. We show that this framework leads to the design of explanatory relations satisfying the rationality postulates of abductive reasoning introduced in [35] and adapted here to the proposed more general framework, and present applications in several logics. This extends previous work on abduction in propositional logics where retraction was defined as a morphological erosion [8–10], as well as abduction in description logics for image understanding [4]. Here weaker properties are required for retraction, that allow defining a larger set of suitable operators for abduction for different logics. This approach is similar to the one proposed for revision in [1], where revision operators were defined from relaxation in satisfaction systems, and then instantiated in various logics. An important feature of the proposed explanations based on retraction is that generation and selection steps are merged, or at least the set of generated hypotheses is reduced, thus facilitating the selection step.

The paper is organized as follows. In Section 2, we recall the useful definitions and properties of satisfaction systems, and provide examples in propositional logic, Horn logic, first order logic, modal propositional logic and description logic. In Section 3 we introduce our first contribution, by defining a notion of cutting, from which explanations are then defined. In Section 4, we propose to define particular cuttings, based on retractions of formulas. Then in Section 5, we instantiate the proposed general framework in various logics.

2. Satisfaction systems

We recall here the basic notions of satisfaction systems needed in this paper. The presentation follows the one in [1], where we give a more complete presentation of satisfaction systems, including the properties and their proofs, that are omitted here.

2.1. Definition and examples

Definition 1 *(Satisfaction system).* A **satisfaction system** $\mathcal{R} = (Sen, Mod, \models)$ consists of

- a set Sen of **sentences**,
- a class Mod of **models**, and
- a satisfaction relation $\models \subseteq Mod \times Sen$.

Let us note that the non-logical vocabulary, so-called *signature*, over which sentences and models are built, is not specified in Definition 1.[1] Actually, it is left implicit. Hence, as we will see in the examples developed in the paper, a satisfaction system always depends on a signature.

Example 1. The following examples of satisfaction systems are of particular importance in computer science and in the remainder of this paper.

Propositional Logic (PL) Given a set of propositional variables Σ, we can define the satisfaction system $\mathcal{R}_\Sigma = (Sen, Mod, \models)$ where Sen is the least set of sentences finitely built over propositional variables in Σ, the symbols $\top$ and $\bot$ (denoting tautologies and antilogies – or contradictions –, respectively), and Boolean connectives in $\{\neg, \vee, \wedge, \Rightarrow\}$, Mod contains all the mappings $\nu : \Sigma \to \{0, 1\}$ (0 and 1 are the usual truth values), and the satisfaction relation $\models$ is the usual propositional satisfaction.

Horn Logic (HCL) A *Horn clause* is a sentence of the form $\Gamma \Rightarrow \alpha$ where Γ is a finite (possibly empty) conjunction of propositional variables and α is a propositional variable. The satisfaction system of Horn clause logic is then defined as for **PL** except that sentences are restricted to be conjunctions of Horn clauses.

Modal Propositional Logic (MPL) Given a set of propositional variables Σ, we can define the satisfaction system $\mathcal{R}_\Sigma = (Sen, Mod, \models)$ where

- Sen is the least set of sentences finitely built over propositional variables in Σ, the symbols $\top$ and $\bot$, Boolean connectives in $\{\neg, \vee, \wedge, \Rightarrow\}$, and modalities in $\{\Box, \Diamond\}$;
- Mod contains all the Kripke models (I, W, R) where I is an index set, $W = (W^i)_{i \in I}$ is a family of functions from Σ to $\{0, 1\}$, and $R \subseteq I \times I$ is an accessibility relation;

[1] The set of logical symbols is defined in each particular logic and does not depend on a theory.