Journal of Applied Logic 13 (2015) 720–769

Contents lists available at ScienceDirect

Journal of Applied Logic

www.elsevier.com/locate/jal

A history based approximate epistemic action theory for efficient postdictive reasoning

Manfred Eppe [a,*], Mehul Bhatt [b]

[a] International Computer Science Institute, Berkeley, USA
[b] University of Bremen, Germany

ARTICLE INFO

Article history:
Accepted 29 June 2015
Available online 28 August 2015

Keywords:
Commonsense reasoning
Action and change
Epistemic reasoning

ABSTRACT

We propose an approximation of the possible worlds semantics ($\mathcal{PWS}$) of knowledge with support for *postdiction* – a fundamental inference pattern for diagnostic reasoning and explanation tasks in a wide range of real-world applications such as cognitive robotics, visual perception for cognitive vision, ambient intelligence and smart environments. We present the formal framework, an operational semantics, and an analysis of soundness and completeness results therefrom.

The advantage of our approach is that only a linear number of state-variables are required to represent an agent's knowledge state. This is achieved by modeling knowledge as the history of a single approximate state, instead of using an exponential number of possible worlds like in Kripke semantics. That is, we add a temporal dimension to the knowledge representation which facilitates efficient postdiction. Since we consider knowledge **h**istories, we call our theory *h-approximation* ($\mathcal{HPX}$).

Due to the linear number of state variables, $\mathcal{HPX}$ features a comparably low computational complexity. Specifically, we show that $\mathcal{HPX}$ can solve the projection problem in polynomial (tractable) time. It can solve planning problems in NP, while e.g. for the action language $\mathcal{A}_k$ [48] this is Σ_2^P-complete. In addition to the temporal dimension of knowledge, our theory supports concurrent acting and sensing, and is in this sense more expressive than existing approximations.

1. Introduction

Commonsense reasoning about action & change is a vibrant research area concerned with developing formal methods for representing and reasoning about high-level knowledge [37,7,10,20]. One particular line of research in reasoning about action & change is that of *epistemic* action theories, e.g. [36,48,47], which are concerned with reasoning about the knowledge of an agent. An agent can form knowledge directly, by executing sensing actions, or by deductive and abductive inference mechanisms.

* Corresponding author.
E-mail addresses: eppe@icsi.berkeley.edu (M. Eppe), bhatt@informatik.uni-bremen.de (M. Bhatt).

http://dx.doi.org/10.1016/j.jal.2015.08.001
1570-8683/

Epistemic action theories are usually represented by a *possible-world* model of knowledge, following the work by Moore [36]. A problem with such approaches is that an exponential number of possible worlds are required to model an agent's knowledge states. Hence, possible worlds-approaches are highly intractable. Approximations with a lower complexity exist (e.g. [48]), but these usually have no support for an important form of inference, called *postdiction* – a kind of logical reasoning that is concerned with inferring knowledge about the past based on present observations (see Examples 1, 2, 3).

Our research addresses the development of approximate epistemic action theories that (i) natively support postdiction, (ii) have a lower complexity bound compared to approaches based on possible worlds, (iii) are provably sound, and (iv) support concurrency. This paper presents the *h-approximation* ($\mathcal{HPX}$), a history-based approximation of $\mathcal{PWS}$ that satisfies criteria (i–iv). This paper mainly elaborates on the theory of the proposed approach. An implementation of the theory in terms of answer set programming [2], along with an empirical evaluation and a case study, are presented in a follow-up paper [14].

Postdictive reasoning

We regard postdiction as a form of reasoning accounting for causal relations between temporally ordered states. Postdiction is abductive reasoning, in the sense that it can be used to explain an observation. However, technically, it can be implemented in a deductive manner, as shown throughout this paper. Within an epistemic action theory, postdictive reasoning can be applied to verify action success and to infer new knowledge about the past. For illustration, consider the following examples:

Example 1 *(The litmus test).* To find out whether a liquid is acidic or alkaline, one can hold a litmus paper into the liquid. If the paper is red, one can postdict that the liquid is acidic, and if the paper is blue, one knows that the liquid is alkaline.

This example, due to Moore [36], illustrates how postdiction is used to determine a world property (the acidic-ness of the liquid) by executing a non-sensing action (holding a paper in the liquid) and observing another world property (the color of the paper) that is a causal consequence of the non-sensing action.

Example 2 *(Grabbing an egg).* Consider a household-robot with an object grabber and a vision system to identify objects. When grabbing a fragile object like an egg, the robot applies a certain force with the grabber and the egg may break. It can postdict that too much force was applied, if it observes that the egg is broken.

The example illustrates how postdiction is the underlying logical inference mechanism for learning action parameters by observing the (non-)success of the action.

Example 3 *(Robot passing through a door).* A robot has the task to open a door and to drive through it. It can execute an open-door command, but it does not have haptic sensors to directly verify the state of the door after executing the command. Hence, considering that the door is sometimes jammed, the robot can never really know whether the door is indeed open or closed after executing the command. However, it does have a location sensor to determine on which side of the door it currently is. If the robot ends up on the other side of the door after driving, then it can postdict that the door was open, and hence that it was not jammed (assuming that the robot knows that the door was initially closed).

This example illustrates an indirect form of postdiction, where a world property (the jammed-ness of the door) is determined by executing a sequence of two non-sensing actions that are causally linked: if the door