Knowledge-Based Systems 27 (2012) 281–301

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

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

Perceptually grounded self-diagnosis and self-repair of domain knowledge

Joshua K. Jones *, Ashok K. Goel

Design & Intelligence Laboratory, School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA 30332, United States

ARTICLE INFO

Article history:
Received 17 March 2011
Received in revised form 10 August 2011
Accepted 17 September 2011
Available online 22 September 2011

Keywords:
Knowledge engineering
Knowledge representation
Symbol grounding
Metareasoning
Compositional classification

ABSTRACT

We view incremental experiential learning in intelligent software agents as progressive agent self-adaptation. When an agent produces an incorrect behavior, then it may reflect on, and thus diagnose and repair, the reasoning and knowledge that produced the incorrect behavior. In particular, we focus on the self-diagnosis and self-repair of an agent's domain knowledge. The core issue that this article addresses is: what kind of metaknowledge may enable the agent to diagnose faults in its domain knowledge? To address this question, we propose a representation that explicitly encodes metaknowledge in the form of *Empirical Verification Procedures* (EVPs). In the proposed knowledge representation, an EVP may be associated with each concept within the agent's domain knowledge. Each EVP explicitly semantically grounds the associated concept in the agent's perception, and can thus be used as a test to determine the validity of knowledge of that concept during diagnosis. We present the empirical evaluation of a system, Augur, that makes use of EVP metaknowledge to adapt its own domain knowledge in the context of a particular subclass of classification problem called Compositional Classification.

1. Introduction

It is generally agreed in AI that the capability of metareasoning, reasoning about reasoning, is essential for achieving human-level intelligence [1–4]. A canonical metareasoning architecture is depicted in Fig. 1. Metareasoning systems extend the basic view of a software agent, where the agent receives percepts from and acts within an environment (called the *object* level in Fig. 1), to include a reflective layer that *monitors* the agent processing and exerts *control* over it, e.g. by altering the object level if it becomes apparent that progress is not being made. Cox [5] and Anderson and Oates [6] review AI research on metareasoning.

In this article, we describe work on enabling metareasoning agents, implemented purely in software, to reflect upon and modify the agent's domain knowledge. Note that this research topic is related to, but distinct from, a more common use of metareasoning for agent adaptation – the adaptation of an agent's processing. We are concerned here specifically with adapting declarative domain knowledge and not an agent's process within the context of a domain. The central question addressed by this research is: what is the form of metaknowledge that will be useful to an agent in reasoning over and adapting its own knowledge? The overarching hypothesis adopted by this work is that *knowledge about domain knowledge (metaknowledge) should be specified in the form of verifiable predictions*.

The next question that arises is: how can one operationalize the verifiable predictions such that the agent can automatically check the correctness of its knowledge? The answer we propose is that each piece of an agent's knowledge may have associated with it procedures consisting of sequences of actions and observations in the environment that can be used to test the veracity of an associated piece of domain knowledge. We call these pieces of metaknowledge Empirical Verification Procedures (EVPs). We hypothesize that they are a form of metaknowledge that will enable a system to successfully self-diagnose and repair domain knowledge. The use of explicitly represented EVP knowledge for the diagnosis and repair of domain knowledge is the specific innovation of the research described here. An interesting implication of this hypothesis is that domain knowledge is grounded in perception, because that knowledge will be considered correct only if it leads to accurate predictions about the world and modified to conform to that ideal of correctness otherwise.

In order to test these hypotheses empirically, we must refine them still further within the context of a specific problem so that we arrive at an implementable level of detail. Since classification is a ubiquitous task in AI [8–11], we have chosen to consider the problem of using metaknowledge for repairing classification knowledge when the classifier supplies an incorrect class label. More specifically, we consider the subclass of classification problems that can be decomposed into a hierarchical set of smaller classification problems; alternatively, problems in which features describing the world are progressively aggregated and abstracted into higher-level abstractions until a class label is produced at the root node. This subclass of classification problems is recognized

* Corresponding author.
E-mail addresses: jkj@cc.gatech.edu (J.K. Jones), goel@cc.gatech.edu (A.K. Goel).

0950-7051/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.knosys.2011.09.012

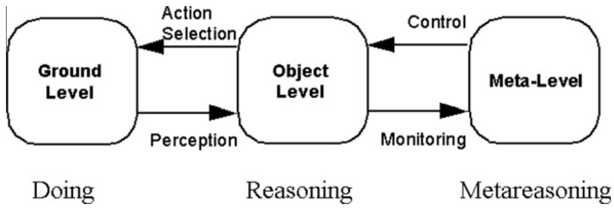

Fig. 1. A basic metareasoning architecture, adapted from [7].

as capturing a common pattern of classification (e.g. [12,13]). In fact, this class of problems is so common that it has been identified as a Generic Task [14]. We will call this classification task *Compositional Classification*, and the hierarchy of abstractions an *Abstraction Network*. In particular, we consider the problem of retrospective, failure-driven adaptation of the content of the intermediate abstractions in the Abstraction Network (and *not* its structure) when the classifier makes an incorrect classification.

Refining our overall hypotheses in the context of Compositional Classification means that intermediate abstractions in the Abstraction Network are chosen such that each abstraction corresponds to a prediction about percepts in the world, metaknowledge comes in the form of verification procedures (EVPs) associated with the abstractions, and metareasoning invokes the appropriate EVPs to perform structural credit assignment [1,15] and then adapt the abstractions. The EVPs explicitly encode the grounding of intermediate abstractions in percepts from the environment, and will be modified when the agent sees evidence that the associated abstraction fails to support appropriate inference at the parent. This architecture for Compositional Classification is depicted in Fig. 2. To support empirical evaluation of our theory within the domain of Compositional Classification, we have implemented a system, Augur, that makes use of EVPs for self-adaptation of Compositional Classification knowledge. In the remainder of this article we illustrate, formalize and evaluate the use of EVPs for self-adaptation of domain knowledge in Abstraction Networks, and present empirical results obtained by applying Augur in both synthetic and real domains.

These hypotheses, and the corresponding observation about the predictive nature of the knowledge used to adapt reasoning processes, suggest an elaboration of the canonical metareasoning architecture of Fig. 1, depicted in Fig. 3. In the view of metareasoning taken in this work, the meta-level detects errors in processing and/or knowledge at the object level based on violations of expectations expressed in terms of the environment. Thus, the meta-level needs to observe not only the object level, but also the ground level. Further, when problems are identified at the object level by this monitoring, the meta-level may need to cause the system to take some actions in the environment in order to gather more information needed to resolve the problems. For example, the meta-level may execute EVPs at intermediate nodes in a classification hierarchy to determine which pieces of knowledge are responsible for an observed top-level classification error. Finally, as shown in Fig. 2, metaknowledge used by the meta-level process may be directly distributed over the object level knowledge structures rather than being strictly confined to separate representations of the meta-level – here, EVPs are encoded as part of an agent's hierarchical classification knowledge.

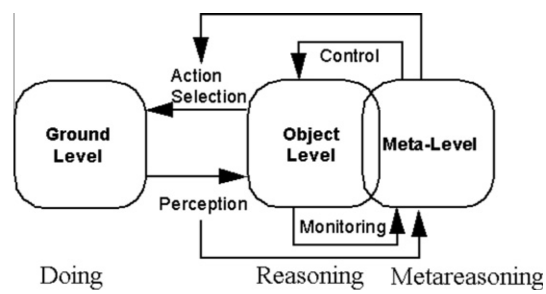

Fig. 3. Elaborated metareasoning architecture.

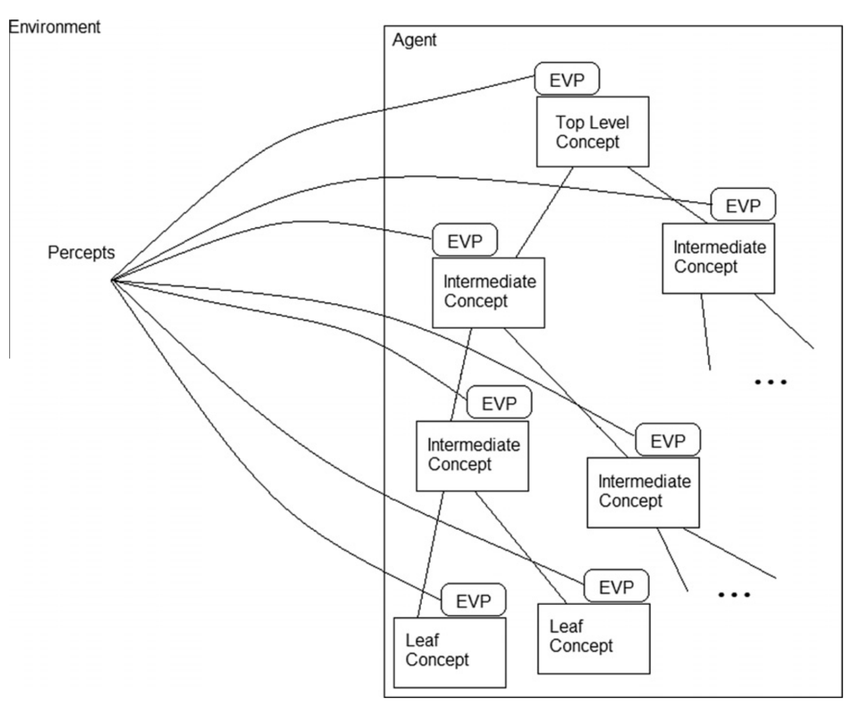

Fig. 2. Hierarchical classification knowledge structure with Empirical Verification Procedures grounding concepts in perception.