ELSEVIER

Artificial Intelligence 97 (1997) 169-193

Artificial Intelligence

Defaults and relevance in model-based reasoning [1]

Roni Khardon [a,*], Dan Roth [b,2]

[a] Aiken Computation Laboratory, Harvard University, Cambridge, MA 02138, USA

[b] Department of Applied Mathematics and Computer Science, Weizmann Institute of Science, Rehovot 76100, Israel

Received September 1995; revised May 1996

Abstract

Reasoning with model-based representations is an intuitive paradigm, which has been shown to be theoretically sound and to possess some computational advantages over reasoning with formula-based representations of knowledge. This paper studies these representations and further substantiates the claim regarding their advantages. In particular, model-based representations are shown to efficiently support reasoning in the presence of varying context information, handle efficiently fragments of Reiter's default logic and provide a useful way to integrate learning with reasoning. Furthermore, these results are closely related to the notion of relevance. The use of relevance information is best exemplified by the filtering process involved in the algorithm developed for reasoning within context. The relation of defaults to relevance is viewed through the notion of context, where the agent has to find plausible context information by using default rules. This view yields efficient algorithms for default reasoning. Finally, it is argued that these results support an incremental view of reasoning in a natural way, and the notion of relevance to the environment, captured by the Learning to Reason framework, is discussed. © 1997 Elsevier Science B.V.

Keywords: Knowledge representation; Common-sense reasoning; Learning to reason; Reasoning with models; Context; Default reasoning

* Corresponding author. Email: roni@das.harvard.edu. Research supported by ARO under grant DAAL03-92-G-0115 and by ONR grant N00014-96-1-0550.

[1] An earlier version of the paper appears in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-95)*.

[2] Email: danr@wisdom.weizmann.ac.il. Research supported by the Feldman Foundation. Part of this work was done while at Harvard University, supported by NSF grant CCR-92-00884, DARPA AFOSR-F4962-92-J-0466 and ONR grant N00014-96-1-0550.

PII S0004-3702(97)00044-1

1. Introduction

A considerable amount of work on the theoretical foundations of artificial intelligence has been devoted to capturing some of the intuitive notions of human reasoning. An introspective view suggests that the notion of *relevance* is central to human reasoning. Many "common-sense" reasoning situations are characterized by the abundance of potentially relevant information sources. Yet, humans seem to pick up just the relevant information and ignore the irrelevant. This ability may account for the speed with which reasoning is performed in everyday situations. In this paper, relevance is viewed as a notion that can be used to reduce the computational cost of reasoning, by focusing on information that pertains to the situation or task at hand, and ignoring the information which does not bear on this situation. We study several tasks in the framework of logical reasoning, and show that relevance information can indeed be useful. In particular, we show that when performing *reasoning with models*—namely, when reasoning is performed by considering examples from the world we reason about—relevance information can be used to efficiently tackle several reasoning tasks.

The generally accepted framework for the study of reasoning in intelligent systems is the knowledge-based system approach. The idea is to store the knowledge in some *representation language* with a well defined meaning assigned to its sentences. The sentences are stored in a knowledge base (*KB*) which is combined with a reasoning mechanism that is used to determine what can be inferred from the sentences in the *KB*. Various knowledge representations can be used to represent the knowledge in a knowledge-based system. Different representation systems (e.g., a set of logical rules, a probabilistic network) are associated with corresponding reasoning mechanisms, each with its own merits and range of applications [14,18]. Given a logical knowledge base, for example, reasoning can be abstracted as a deduction task: determine whether a sentence, assumed to capture the situation at hand, is logically implied by the knowledge base.

It is also widely agreed that a large part of our everyday reasoning involves arriving at conclusions that are not entailed by our "theory" of the world. Many conclusions are derived in the absence of information that is sufficient to imply them. This type of reasoning is naturally nonmonotonic since further evidence may force us to revise our conclusions. Several formalizations trying to capture this situation have been studied, and of particular interest to us here are theories for reasoning with "defaults" (see e.g. [20]). In this approach, the true knowledge about the world is augmented by a set of *default rules* that are meant to capture "typical" cases. The quest is for a reasoning system that, given a query, responds in a way that agrees with what we know about the world and (some of) the default assumptions, and at the same time supports our intuition about a plausible conclusion.

Computational considerations, however, render this approach inadequate for common-sense reasoning. This is true not only for the task of deduction, but also for many other forms of reasoning that have been developed. All those were shown to be even harder to compute than the original formulation [23,25]. This holds in particular for various formalizations of default reasoning [6,17,26], where the increase in complexity is clearly at odds with the intuition that reasoning with defaults should somehow reduce the complexity of reasoning. This remains true, even when we severely restrict the expressiveness of the knowledge base, the default rules and the queries allowed. In this paper we show that model-based representations can be used to overcome some of the above-mentioned difficulties, and that the notion of relevance is useful in deriving these results.

We incorporate the notion of relevance into the study of reasoning by introducing the task of reasoning within context. It has been argued that in real life situations, one normally completes a lot of missing context information when answering queries [12]. We model this situation by augmenting the agent's knowledge about the world with context-specific information. Reasoning within context is therefore a deduction task, where some additional constraining information is added to the knowledge base. We formalize this task as the problem of *reasoning within a varying context.* The intuition is that the availability of additional context information should make the reasoning task easier, by restricting the domain one needs to reason about. As we show, if the agent performs model-based reasoning then context information can be easily used, and yields efficient reasoning.

In model-based reasoning [5,9] the knowledge base is represented as a set of models (satisfying assignments, examples) of the world rather than a logical formula describing it. When a query is presented, reasoning is performed by evaluating the query on these models. It is not hard to motivate a model-based approach to reasoning from a cognitive point of view and indeed most of the proponents of this approach have been cognitive psychologists [3,4,11], who have alluded to the notion of "reasoning from examples" on a qualitative basis. In the AI community this approach can be seen as an example of Levesque's notion of "vivid" reasoning [12,13], and is somewhat related to Minsky's frames-theory [15] and to some of the work in case-based reasoning [10].

Given a model-based representation of the knowledge base *KB*, and a query α, the task of deciding whether *KB* implies α (denoted $KB \models \alpha$) can be performed in a straightforward way: Evaluate α on all the models in the representation. If you find a model of *KB* which does not satisfy α, then $KB \not\models \alpha$, otherwise conclude that $KB \models \alpha$. Clearly, if the model-based representation contains all the models of *KB* then, by definition, this approach verifies the implication, and yields correct deduction. But representing *KB* by explicitly holding *all* the possible models is not plausible. A model-based approach becomes feasible if *KB* can be replaced by a small model-based representation and still support correct deduction.

The theory of model-based representations developed in [9] (generalizing the theory developed in [5] for the case of Horn expressions) characterizes the propositional languages for which model-based representations support efficient deduction and abduction. It is shown that in many cases in which deduction is NP-hard in the formula-based setting, the model-based representation is small (polynomial in the number of propositional variables in the domain). Thus, in the model-based setting, correct and efficient reasoning can be obtained in cases where such algorithm were not known before.

When reasoning within context, our general knowledge about the world is combined with additional constraints, those that are relevant to the current situation. This task has an easy and natural implementation when using model-based representations. The algorithm simply filters out models in the representation which are not relevant to the current context, namely models that are not consistent with the context information. The remaining models are used as before for reasoning with models. The filtering process can be performed as a background process whenever the context changes, thereby speeding up the reasoning. We characterize several propositional languages for which this simple algorithm works correctly and efficiently. We note that reasoning with models is essential for providing the computational advantage when reasoning within context. The filtering algorithm cannot be performed when reasoning with formulas, and additional context information does not necessarily make the task easier.

Intuitively, the task of default reasoning also aims at capturing some notion of relevance in that it enforces additional constraints to weed out non-relevant cases and focus on the typical cases. In *default reasoning* [19] an agent is given a representation of the world, and a set of (sometimes conflicting) default rules, and has to assess whether a query q can be concluded "by default", namely using the standard knowledge and the default rules. We show that default reasoning is a generalization of reasoning within context, in which the reasoner has many context rules, which may be conflicting. Namely, default rules do not express explicitly what the relevant context information is, but rather implicitly capture all plausible scenarios. While non-typical cases can be ignored, a new source of computational difficulty is introduced when default rules are not compatible with each other. In this case, the agent can only use a subset of these rules for deriving its conclusions, and there may be many possibilities for choosing such subsets. To reach a conclusion the agent must enumerate the various possible "contexts" (called extensions in default reasoning), and perform the reasoning relative to one context at a time. In some sense, the need to search through possible contexts in order to derive a conclusion is at odds with the notion of relevance and is the source of the additional computational difficulty in default reasoning.

Nevertheless, we show that in some cases model-based representations provide a way out of the computational difficulty. In particular, in these cases, model-based representations capture all possible contexts in an accessible form. Thus, one can enumerate the contexts, and then use a subroutine for reasoning within context to perform the inference task. We give efficient algorithms for (both credulous and skeptical) default reasoning tasks, for several classes of world knowledge, default rules and queries. We also show that in some cases the task of diagnosis can be performed by similar techniques. Our results provide efficient solutions for cases that are not known to be solvable if the knowledge base is represented with formulas.

Our notion of relevance as a way to reduce the computational cost of reasoning can be also used at a higher level, that of *relevance for the task.* While above, relevance or context information was used to prune the knowledge representation for a particular situation, it is also possible to prune a representation relative to a general task. For example, if it is known that all the queries presented to the agent come from some restricted language then an approximation of *KB* relative to this language is sufficient in order to support correct reasoning with these queries. This idea can be formalized using the notion of least upper bound approximations [9,27]. In fact, the model-based representations used for reasoning within context and default reasoning capture such approximations, and thus use this notion of relevance as well. Here again, the model-based representations are essential since formula-based representations of these approximations do not support efficient reasoning.

Our results show that knowledge, which holds within a specific context and is available in a form of a model-based representation, can be used to reason correctly within this context. Therefore, our treatment of reasoning within context supports the view that an intelligent agent can construct a representation of the world incrementally by pasting together many "narrower" views from different contexts. This suggests the notion of *relevance to the environment*. Namely, the agent can incrementally construct a representation that is relevant to its environment, in the sense that it supports correct reasoning there. This intuitive idea is formalized in a more general setting in the Learning to Reason framework [7], and is discussed also in [8,22,31]. We discuss two results in the Learning to Reason framework that use model-based representations in order to exploit the relevant information in the reasoning process.

To summarize, this paper studies reasoning with model-based representations, and further substantiates the claim regarding their computational advantages. In particular, model-based representations are shown to efficiently support reasoning in the presence of varying context information, handle efficiently fragments of Reiter's default logic and provide a useful way to integrate learning with reasoning. On a more philosophical level, we suggest that it is useful to view these results through the notion of relevance: the computational cost of reasoning can be reduced by using only the information that is relevant to the current situation, the general task being performed, or the environment. We discuss several aspects of relevance, and show that their use is enabled by model-based representations. In particular:

(1) The use of relevance information is best exemplified by the filtering process involved in the algorithm for reasoning within context.
(2) The relation of defaults to relevance can be viewed through the notion of context, resulting in efficient algorithms for default reasoning.
(3) Relevance information can also be used at a level of a global task by a priori restricting the attention to information sufficient for the task.
(4) The Learning to Reason framework allows for incrementally constructing a knowledge representation that is relevant for the environment.

The rest of the paper is organized as follows: In Section 2 we introduce some definitions and the notation used throughout the paper. In Section 3 we briefly present some results from the theory of reasoning with models. In Section 4 we discuss the task of reasoning within context and present an efficient model-based algorithm for it. In Section 5 we discuss default reasoning with models and an application to diagnosis. In Section 6 we discuss knowledge approximations as capturing information relevant for a task. In Section 7 we briefly present the Learning to Reason framework and its relation to relevance, and Section 8 concludes with a summary.

2. Preliminaries

We consider problems of reasoning where the "world" is modeled as a Boolean function $W : \{0,1\}^n \to \{0,1\}$. We use interchangeably the terms propositional expression and Boolean function, and likewise for propositional language and a class of Boolean functions. We denote classes of Boolean functions by $\mathcal{F}$, $\mathcal{G}$, and functions by f, g.

Let $X = \{x_1, \ldots, x_n\}$ be a set of *variables*, each of which is associated with a world's attribute and can take the value 1 or 0 to indicate whether the associated attribute is true or false in the world. *Assignments* are mappings from X to $\{0, 1\}$, and we treat them as elements in $\{0, 1\}^n$ with the natural mapping. Assignments in $\{0, 1\}^n$ are denoted by x, y, z, and $weight(x)$ denotes the number of 1 bits in the assignment x. A *clause* is a disjunction of literals, and a CNF formula is a conjunction of clauses. For example, $(x_1 \vee \overline{x_2}) \wedge (x_3 \vee \overline{x_1} \vee x_4)$ is a CNF formula with two clauses. A *term* is a conjunction of literals, and a DNF formula is a disjunction of terms. For example, $(x_1 \wedge \overline{x_2}) \vee (x_3 \wedge \overline{x_1} \wedge x_4)$ is a DNF formula with two terms. A CNF formula is *monotone* if all the literals in it are positive (unnegated). A CNF formula is Horn if every clause in it has at most one positive literal. A CNF formula is k-quasi-Horn if there are at most k positive literals in each clause. It is a k-quasi-reversed-Horn if there are at most k negative literals in each clause. A DNF formula is k-quasi-monotone DNF if there are at most k negative literals in each term.

Every Boolean function has many possible representations and, in particular, both a CNF representation and a DNF representation. By the DNF size of f, denoted $|DNF(f)|$, we mean the minimum number of terms in any DNF representation of f. Similarly, the CNF size of f, denoted $|CNF(f)|$, is the minimum number of clauses in any CNF representation of f.

An assignment $x \in \{0, 1\}^n$ satisfies f if $f(x) = 1$; such an assignment x is called a *model* of f. If f is a theory of the "world", a satisfying assignment of f is sometimes referred to in the literature as a possible world. By "f implies g", denoted $f \models g$, we mean that every model of f is also a model of g. Throughout the paper, when no confusion can arise, we identify a Boolean function f with the set of its models, namely $f^{-1}(1)$. Observe that the connective "implies" ($\models$) used between Boolean functions is equivalent to the connective "subset or equal" ($\subseteq$) used for subsets of $\{0, 1\}^n$. That is, $f \models g$ if and only if $f \subseteq g$.

3. Reasoning with models

In this section we briefly present some results from the monotone theory of Boolean functions [2] and the theory of reasoning with models [5,9]. All the results in this section have appeared elsewhere. For a detailed discussion see [9].

Consider a propositional knowledge base W and let α be a propositional query. The model-based strategy for the deduction problem $W \models \alpha$ is to try and verify the implication relation using model evaluation. Fig. 1 describes the algorithm MBR, which uses a set of models Γ as a knowledge base. When presented with a query α the algorithm evaluates α on all the models in Γ. If a counterexample x is found such that $\alpha(x) = 0$, then the algorithm returns "No". Otherwise it returns "Yes".

Clearly, the model-based approach solves the inference problem if Γ is the set of *all* models (satisfying assignments) of W. However, the set of all models might be too large, making this procedure computationally infeasible. A model-based approach

Algorithm MBR(Γ, α):

Test set: A set $\Gamma \subseteq W$ of possible assignments.
Test: If there is an element $x \in \Gamma$ which does not satisfy α, return "No". Otherwise, return "Yes".

Fig. 1. MBR: model-based reasoning.

becomes useful if one can show that it is possible to use a fairly small set of models as the test set, and still perform reasonably good inference.

In the rest of this section we describe general conditions under which this can be done. An example of the technical notions presented here is given at the end of the section.

3.1. Monotone theory

Definition 1 (*Order*). We denote by $\leqslant$ the *usual partial order* on the lattice $\{0,1\}^n$, the one induced by the order $0 < 1$. That is, for $x, y \in \{0,1\}^n$, $x \leqslant y$ if and only if $\forall i$, $x_i \leqslant y_i$. For an assignment $b \in \{0,1\}^n$ we define $x \leqslant_b y$ if and only if $x \oplus b \leqslant y \oplus b$ (Here $\oplus$ is the bitwise addition modulo 2.) We say that $x > y$ if and only if $x \geqslant y$ and $x \neq y$.

Intuitively, if $b_i = 0$ then the order relation on the ith bit is the normal order; if $b_i = 1$, the order relation is reversed, that is, $1 <_{b_i} 0$.

The *monotone extension of* $z \in \{0,1\}^n$ with respect to b is defined as

$$\mathcal{M}_b(z) = \{x \mid x \geqslant_b z\}.$$

The *monotone extension of* f with respect to b is defined as

$$\mathcal{M}_b(f) = \{x \mid x \geqslant_b z, \text{ for some } z \in f\}.$$

Notice that throughout we treat the function f as the set of its satisfying assignments, and therefore the above notation is natural. The set of *minimal assignments of* f with respect to b is defined as

$$\min_b(f) = \{z \mid z \in f, \text{ such that } \forall y \in f,\ z \not>_b y\}.$$

Definition 2 (*Basis*). A set B is a *basis for* f if $f = \bigwedge_{b \in B} \mathcal{M}_b(f)$. B is a basis for a class of functions $\mathcal{F}$ if it is a basis for all the functions in $\mathcal{F}$.

The importance of these definitions [2] is that every Boolean function has a basis B, and can be represented as follows:

$$f = \bigwedge_{b \in B} \mathcal{M}_b(f) = \bigwedge_{b \in B} \bigvee_{z \in \min_b(f)} \mathcal{M}_b(z). \tag{1}$$

This representation yields a necessary and sufficient condition describing when $x \in \{0,1\}^n$ satisfies f:

Corollary 3. *Let B be a basis for f, and $x \in \{0,1\}^n$. Then, $f(x) = 1$ if and only if for every basis element $b \in B$ there exists $z \in \min_b(f)$ such that $x \geqslant_b z$.*

It is known [2] that for every b, the size of $\min_b(f)$ is bounded by the size of its DNF representation. Further, a set of assignments which falsify every clause in a CNF representation of f is a basis for f. Therefore, f has a basis whose size is bounded by $|CNF(f)|$. Some important function classes have a small *fixed* basis, irrespective of the CNF size of the function:

- *Horn formulas*: A basis for this class is $B_H = \{u \in \{0,1\}^n \mid weight(u) \geqslant n-1\}$, since every Horn clause is falsified by an assignment in B_H. Clearly, $|B_H| = n+1$.
- *k-quasi-Horn formulas*: $B_{H_k} = \{u \in \{0,1\}^n \mid weight(u) \geqslant n-k\}$ is a basis for this class. Clearly, $|B_{H_k}| = O(n^k)$. Similarly, there is a basis for k-quasi-reversed-Horn formulas.
- $\log n$*-CNF formulas*: A Boolean function with a CNF representation in which the clauses contain at most $O(\log n)$ literals. A basis for this class is derived using a combinatorial construction called an (n,k)-universal set. An (n,k)-universal set is a set of assignments $\{a_1, \ldots a_t\} \subseteq \{0,1\}^n$ such that every subset of k variables assumes all of its 2^k possible assignments in the a_i's. It is easy to see that an (n,k)-universal set includes a falsifying assignment for any clause of length k and therefore it forms a basis for the class of k-CNF formulas. It is known [1, 16] that for $k = \log n$ one can construct $(n, \log n)$-universal sets of size $O(n^3)$ and therefore $|B_{\log n\text{-CNF}}| = O(n^3)$.
- *Common queries*: A function is *common* if every clause in its CNF representation is taken from one of the above classes. The union of the bases for these classes is a basis, B_C, for all common functions. We refer to this class as the class of common queries.

3.2. Deduction

We can now characterize a model-based knowledge base for which the algorithm MBR is successful.

Definition 4. For a function $f \in \mathcal{F}$, the set $\Gamma = \Gamma_f^B$ of *characteristic models* of f is the set of all minimal assignments of f with respect to a set $B \subset \{0,1\}^n$. Formally,

$$\Gamma_f^B = \bigcup_{b \in B} \{z \in \min_b(f)\}.$$

If, in addition, B is a basis for a class $\mathcal{G}$ of functions, then we say that Γ_f^B is the *model-based representation* of f with respect to queries in $\mathcal{G}$. The following theorems describe the basic result of the theory of reasoning with models, its application to common queries, and a bound on the size of the model-based representation.

Theorem 5 (Khardon and Roth [9]). *Let f be any Boolean function, and $\alpha \in \mathcal{G}$, where B be a basis for $\mathcal{G}$. Then $f \models \alpha$ if and only if for every $u \in \Gamma_f^B$, $\alpha(u) = 1$. That is, model-based deduction using Γ_f^B, is correct.*

Corollary 6 (Khardon and Roth [9]). *Let f be any Boolean function. Then for any common query α, $f \models \alpha$ if and only if for every $u \in \Gamma_f^{B_C}$, $\alpha(u) = 1$. That is, model-based deduction using $\Gamma_f^{B_C}$, is correct.*

Theorem 7 (Bshouty [2]). *Let f be any Boolean function, and B a basis. Then, the size of the model-based representation of f is*

$$|\Gamma_f^B| \leqslant \sum_{b \in B} |\min_b(f)| \leqslant |B| \cdot |DNF(f)|.$$

We note that this bound is tight in the sense that for some functions the size of the DNF is indeed required. It does however allow for an exponential gap in other cases. Namely, there are functions with an exponential-size DNF and a linear-size model-based representation [9]. It is also interesting to compare the size of this representation to the size of other representations for functions. Examples in [5] show that there are cases where the (Horn CNF) formula representation is small and the model-based representation is exponentially large, and vice versa. For a discussion of these issues, as well as other properties of characteristic models, see [9].

Example. Let f have the CNF representation

$$f = (x_1 \vee x_2 \vee x_3) \wedge (x_1 \vee x_2 \vee x_4) \wedge (\overline{x_1} \vee \overline{x_2} \vee x_3 \vee \overline{x_4}).$$

The function f has 12 (out of the 16 possible) satisfying assignments. The non-satisfying assignments of f are:[3] $\{0000, 0001, 0010, 1101\}$.

If we want to be able to answer all possible Horn queries with respect to f we need to use the Horn basis $B_H = \{1111, 1110, 1101, 1011, 0111\}$. Each of the models 1111, 0111, 1011, 1110 satisfies f and therefore for each of these, $\min_b(f) = \{b\}$. For $b = 1101$, the minimal elements can be found by drawing the corresponding lattice and checking which of the satisfying assignments of f are minimal. This yields $\min_{1101}(f) = \{1100, 1111, 1001, 0101\}$. We therefore get that $\Gamma_f^{B_H} = \{1111, 0111, 1011, 1100, 1001, 0101, 1110\}$. Note that it includes only 7 out of the 12 satisfying assignments of f.

Clearly, in general $\Gamma_f^{B_H} \subseteq f$, and therefore model-based deduction never makes mistakes on queries that are implied by f. Furthermore, for the Horn query $\alpha_1 = x_1 \wedge x_3 \rightarrow x_2$, reasoning with $\Gamma_f^{B_H}$ will find the counterexample 1011 and deduce correctly that $f \not\models \alpha_1$.

We note that in general, if f is given in its CNF representation, solving the problem $f \models \alpha$ is co-NP-complete, even when α is a Horn query.

[3] An element of $\{0,1\}^n$ denotes an assignment to the variables $x_1, \ldots, x_n$ (i.e., 0011 means $x_1 = x_2 = 0$, and $x_3 = x_4 = 1$).

4. Reasoning within context

It has been argued that in real life situations, one normally completes a lot of missing "context" information when answering queries [12]. For example, if asked at a conference how long it takes to drive to the airport, we would probably assume (unless specified otherwise) that the question refers to the city in which the conference is held, rather than to the place where we live (and have been to the airport more times). This corresponds to assigning the value "true" to the attribute "here" for the purpose of answering the question. Sometimes we need a more expressive language to describe our assumptions regarding the current context and assume, say, that some rule applies [26]. For example, we may assume (in the "conference" context) that if someone has a car, then it is a rental car. Thus, reasoning within context may be viewed as a deduction task, where some additional constraining information is added to the knowledge base. Our use of context is closely related to the notion of relevance, since one can use context information in order to concentrate on the relevant knowledge and ignore the irrelevant. Intuitively, this should also make the reasoning task easier. Indeed, as we show, this holds in a formal sense when using model-based reasoning.

Let W be a Boolean function that describes our knowledge about the world. A "first principle" way to formalize the above intuition is the following: we want to deduce a query α from W, if α can be inferred from W given that the query refers to the current context. Namely, the instances of W which are relevant to the query must also satisfy the context condition d, a conjunction of some literals and rules. We denote this question by $W \models_d \alpha$.

Notice that it is possible that $W \models_d \alpha$ but $W \not\models \alpha$, if all the satisfying assignments of W that do not satisfy α do not satisfy d. Formalized this way, the problem $W \models_d \alpha$ is equivalent to the problem $W \wedge d \models \alpha$. Thus, a theorem proving approach to reasoning does not necessarily provide any computational advantage in solving this reasoning problem.

Let $W \in \mathcal{F}$, $\alpha \in \mathcal{G}$ and let B be a basis for $\mathcal{G}$. From Theorem 5 it is clear that given $\Gamma^B_{W\wedge d}$, the set of characteristic models for $W \wedge d$, model-based reasoning can be used to solve the reasoning problem $W \wedge d \models \alpha$. However, we consider here a more general problem: given Γ^B_W we are interested in performing inference according to $\models_d$ with it, where the "context condition" d may vary.

From our model-theoretic definition of the connective $\models_d$ it is clear that if one has *all* the models of W then, by filtering out all the models that do not satisfy d and performing the model-based test on the remaining models, one answers $W \wedge d \models \alpha$ correctly. The algorithm C-MBR, presented in Fig. 2 does just that, with the set Γ.

The following theorems show that, under some conditions, a compact model-based representation behaves like the complete set of models of a theory. Namely, the filtering algorithm C-MBR provides correct reasoning.

Theorem 8. *Given Γ^B_W, the algorithm C-MBR correctly solves the reasoning problem $W \models_d \alpha$ for every d such that B is a basis for $d \to \alpha$.*

Algorithm C-MBR(Γ, d, α):

Test set: Consider only those elements of Γ which satisfy d.
Test: If there is such an element which does not satisfy α, return "No".
Otherwise, return "Yes".

Fig. 2. C-MBR: model-based reasoning within context.

Proof. Clearly, $W \models_d \alpha \equiv W \wedge d \models \alpha \equiv W \models \bar{d} \vee \alpha \equiv W \models (d \rightarrow \alpha)$. Therefore, from Theorem 5, when B is a basis for $d \rightarrow \alpha$, Γ_W^B can be used for model-based reasoning with it. Models of W that do not satisfy d are useless as counterexamples since $d \rightarrow \alpha$ always holds and therefore, the test set of Algorithm C-MBR produces the correct answer. □

Theorem 9. *The following conditions on α, B and d guarantee that C-MBR supports correct reasoning within context.*

(i) *Let α be a k-quasi-Horn query and $B = B_{\mathrm{H}_{k+r}}$, a basis for $(k+r)$-quasi-Horn theories. If d is a Boolean function that can be represented as a r-quasi-monotone DNF then B is a basis for $d \rightarrow \alpha$. In particular, when $k = 1$ and $r = 0$, this holds for any Horn formula α and any monotone Boolean function d.*

(ii) *Let α be a $\log n$-CNF query and B a basis for $2\log n$-CNF theories. If d is a conjunction of up to $\log n$ arbitrary rules (disjunctions) then B is a basis for $d \rightarrow \alpha$.*

Proof. Assume first that d is given as a DNF expression $\vee_{i\in I} t_i$ and α is given as a CNF expression $\wedge_{j\in J} c_j$. In this case,

$$\begin{aligned} d \rightarrow \alpha &= \Big(\bigvee_{i\in I} t_i\Big) \rightarrow \Big(\bigwedge_{j\in J} c_j\Big) \\ &= \Big(\bigwedge_{i\in I} \bar{t}_i\Big) \vee \Big(\bigwedge_{j\in J} c_j\Big) \\ &= \bigwedge_{i\in I, j\in J} (\bar{t}_i \vee c_j). \end{aligned}$$

For case (i), consider first the case $k = 1, r = 0$. Since d is monotone, every term in a DNF expression for d is monotone, and therefore $(\bar{t}_i \vee c_j)$ is a Horn disjunction. In general, every term in d can contribute at most r positive literals, and c_j can contribute at most k positive literals to $(\bar{t}_i \vee c_j)$. Therefore, $(\bar{t}_i \vee c_j)$ is in $(k+r)$-quasi-Horn. For case (ii), d is a CNF expression with at most $\log n$ clauses, and therefore can be written as a DNF expression in which every term has at most $\log n$ literals. Therefore $(\bar{t}_i \vee c_j)$ has at most $2\log n$ literals.

Finally, notice that we do not need to get d as a DNF expression. The analysis uses this expression to show that $d \rightarrow \alpha$ belongs to the desired class, but the algorithm evaluates d and α directly, using the representation given to it and the filtering algorithm. □

It is interesting to note that the size of the expression for $d \rightarrow \alpha$ described in the above proof might be exponentially large. However, it appears only in the analysis. We do not actually compute this expression in the algorithm. Rather, filtering examples according to d is sufficient.

Notice that our definition of characteristic models, and ultimately also the set of filtered models used in a particular context, depends on the basis B. This may seem confusing at first since a class of queries may have more than one possible basis and the choice of B is arbitrary. However, note that *any* basis for the class of queries can be used to represent *all* the queries in this class, and therefore a set of characteristic models that is based on it captures all the information needed to reason with the queries. This is similar to the situation that arises when representing knowledge using formulas; the same Boolean function can be represented in many ways, and even using several different representations from the same class (for example, in general there is no unique minimal representation for Boolean functions in CNF form).

The approach presented in this section can be viewed as a process of *augmenting* a model-based representation Γ with a set of rules. Given a model-based representation Γ_W^B of W, any rule that holds in W cannot help in answering queries, since it does not filter out any assignment of W, and is thus redundant. However, the context rules do not hold in W and thus augmenting W with them modifies the set of conclusions. As we have shown, in order to reason within context, we need to maintain a model-based representation with respect to a basis that is slightly larger than the basis in the pure deductive case.

4.1. Context and relevance

Our treatment of context information appeals to the intuitive notion of relevance. The algorithm C-MBR uses the context information d to filter out the irrelevant information in the knowledge base. The irrelevant information in this case is the set of models that do not correspond to the current context. The use of a model-based representation facilitates efficient filtering, since it only requires evaluating the current context on the elements of the representation. A natural approach would be to perform the filtering algorithm in the background, whenever the context changes. This way, the reasoning itself will be faster since at any time only evaluation on the models of the current context is required.

This should be contrasted with a formula-based representation, where adding context information does not necessarily help the computational task. There, the formula d is conjuncted with the formula W, and theorem proving is used. Since $W \wedge d$ is not necessarily simpler to reason with than W, the task, in general, does not become easier.

5. Default reasoning with models

Default reasoning is a formal framework for arguing about typical cases. A default rule captures the idea that under normal circumstances we may assume that certain conditions, reminiscent of context information, are satisfied. In default reasoning, however, one might have many rules whose conclusions may be conflicting. As a result the correct context, that combines these conclusions, is not obvious from the set of rules and there may be more than one option for such context. This suggests that default reasoning can be viewed as a generalization of reasoning within context where, given a query, the agent has to search for a correct context before reasoning. Intuitively, default reasoning is, at least partly, aimed at capturing some notion of relevance, by ignoring the non-typical (irrelevant) cases. Therefore, the use of defaults is expected to make the reasoning problem easier. However, default reasoning is actually harder than deductive reasoning, and the computational difficulties can be traced to the above-mentioned search for the correct context. Nevertheless, as our results show, in some cases model-based representations can capture all the possible contexts in an accessible form, and thus by enumerating contexts, or concentrating on one context at a time, an efficient solution can be found.

We will concentrate here on a special case of Reiter's default logic [19], applied to propositional logic.

In Reiter's default logic, default rules have the form $\frac{\alpha:\beta}{\gamma}$, which should read as "if α holds and it is consistent to assume β then conclude γ". The case with $\beta = \gamma$ is called *normal defaults*, and α is called a prerequisite. The discussion below considers normal defaults with empty prerequisites, denoted by $\frac{:\beta}{\beta}$. In this case, we denote by D the set of Boolean functions β, and say that D is the set of default rules. We sometimes treat a collection of rules as their conjunction. That is, $D(x) = 1$ means $\bigwedge_{d\in D} d(x) = 1$.

Definition 10. For normal defaults with empty prerequisites $\frac{:\beta}{\beta}$ we define: A default rule is *simple* if β is a single literal. The rule is *positive* if β is any monotone function. The rule is *positive simple* if β is a positive literal.

Notice that the theory for diagnosis [21] and the closed world defaults [19] can be described using simple defaults.

A *default theory* is a pair (D, W) where D is a set of default rules, and W is a propositional expression. An *extension* of (D, W) is defined using a fixed point operator [19]. For our special case the following theorem gives an alternative and simpler definition: (The operator $Th(R)$ denotes the theorem closure of R.)

Theorem 11 (Reiter [21, p. 88]). *Let D be a set of normal defaults with empty prerequisites. E is an extension of (D, W) if and only if $E = Th(W \wedge S)$, for some maximal subset*[4] *S of D such that $W \wedge S$ is consistent.*

Using this theorem as the definition for extension we can identify a maximal consistent subset S with each extension E. We denote this subset by S_E. Since S_E is consistent with W, we get that an extension E includes q if and only if $W \wedge S_E \models q$. In the original formalization [19] a knowledge base is defined to imply a query if there is

[4] To avoid confusion we emphasize that S and D are considered as sets of rules. S is maximal in the sense that no additional rule can be added to it while preserving consistency with W. Thus, if $S_1 \subset S_2$, the Boolean function S_2 is the conjunction of more rules and therefore $S_2 \models S_1$.

an extension in which the query holds. Following [6,29] we call this task *credulous* default reasoning; the case in which all extensions are taken into consideration is called *skeptical* default reasoning.

Formally, the *credulous default reasoning task* $CDEF(D,W,q)$ is defined as follows: given a default theory (D,W) and a propositional expression q, decide whether there exists an extension E of (D,W) such that $q \in E$.

The *skeptical default reasoning task* $SDEF(D,W,q)$ is defined as follows: given a default theory (D,W) and a propositional expression q, decide whether for all the extensions E of (D,W), $q \in E$.

Clearly, if W is consistent with the set of all rules in D, then there is only one maximal consistent subset S of D, the one which contains all these rules. In this case both credulous and skeptical default reasoning reduce to reasoning within the context D, as discussed earlier. The main difficulty which arises in the general case is that W may not be consistent with all of D.

Next we present positive results on default reasoning using a model-based representation. As in the case of deductive reasoning [9], the efficient results we present hold in cases where there is no known efficient solution when reasoning with formulas. The exact complexity relation is somewhat more subtle than in the deductive case. There, efficient solutions where presented for problems that are NP-hard (under randomized reductions) given a formula-based representation. In the current case, the default reasoning task is NP-hard [28] when the knowledge base is Horn, all the default rules are positive literals, and the query is a single positive literal. Our results provide an algorithm for this class of problems, which is polynomial in the size of the model-based representation. This representation, though, may be exponential in the size of the Horn expression, as is the case, for example, for the problems used in the reduction in [28]. Thus, strictly speaking, we do not prove an advantage in this special case. Nevertheless, our results provide efficient algorithms in cases where they were not known to exist before.

We present two algorithms, *CD*-MBR and *SD*-MBR, which handle the credulous and skeptical default reasoning tasks, respectively. Both algorithms are similar to the abduction algorithm[5] developed in [5] and used in [9].

5.1. Credulous default reasoning

We start by describing the algorithm *CD*-MBR, which is presented in Fig. 3. Let $\Gamma = \Gamma_W$ be a model-based representation of W. (The monotone basis will be defined later.) The algorithm *CD*-MBR receives Γ, D and a query q as input. It starts by enumerating all the models in Γ. When it finds a model z in which the query holds (i.e., $q(z) = 1$), it sets S to be the set of all the rules in D that this model satisfies. The algorithm then tests whether $W \wedge S \models q$ by calling the procedure C-MBR to decide whether $W \models_S q$. If the answer is "Yes" the algorithm returns "Yes"; otherwise, it continues to test the next model in Γ. If all the models in Γ have been scanned and no good extension has been found the algorithm says "No".

[5] Our results were inspired by the connections between abduction and default reasoning developed in [25].

Algorithm *CD*-MBR(Γ, D, q):

Do for all models $z \in \Gamma$ such that $q(z) = 1$
  Let $S = \{d \in D \mid d(z) = 1\}$
  If C-MBR(Γ, S, q) answers "Yes", return "Yes".
EndDo
Return "No". /* No extension found */

Fig. 3. *CD*-MBR: default reasoning with a model-based representation.

Assume the algorithm is run with a model-based representation $\Gamma = \Gamma_W^B$ and the query presented to it is q. The following two conditions on B and D are used to characterize the cases in which the algorithm is successful.

Condition 12. B is a basis for $S \to q$, for all $S \subseteq D$.

Condition 13. For all $S \subseteq D$, for all u such that $S(u) = 1$, there is a prime implicant t of S, and a basis element $b \in B$, such that $t(u) = t(b) = 1$.

Theorem 14. *Given Γ_W^B, the algorithm CD-MBR solves the credulous default reasoning task CDEF(D, W, q) correctly, whenever Condition* 12 *and Condition* 13 *hold.*

Proof. We need to prove:
(i) if the algorithm returns "Yes" then the desired extension exists, and
(ii) if there is an extension that contains q, then the algorithm returns "Yes".

For (i), since Condition 12 holds, Theorem 8 implies that C-MBR is correct, that is, $W \wedge S \models q$. By construction, S is a subset of D for which $W \wedge S$ is consistent. If $S^* \supseteq S$ is a maximal subset of D (with respect to the property of consistency with W), then clearly $W \wedge S^* \models W \wedge S \models q$, and the required extension exists.

For (ii), assume that there is an extension E that contains q. By definition, the existence of E implies that there exists a subset $S_E \subseteq D$ such that $W \wedge S_E$ is consistent and, $W \wedge S_E \models q$. Thus, there is an assignment $u \in W$ such that $S_E(u) = 1$ and therefore also $q(u) = 1$.

Condition 13 implies that there exists a prime implicant t of S_E and a basis element $b \in B$ such that $S_E(u) = t(u) = t(b) = 1$. Thus, u, and b agree on all the literals that appear in t, and therefore for all z such that $z \leqslant_b u$, $t(z) = S_E(z) = 1$. Thus, if $w \in \Gamma_W^B$ is such that $w \leqslant_b u$ (and since $u \in W$, w like that always exists), then $S_E(w) = 1$ and since $W \wedge S_E \models q$ also $q(w) = 1$.

Now, consider the set S that the algorithm *CD*-MBR uses in the iteration for $w \in \Gamma_W^B$. By construction the algorithm will compute a set which is identical to S_E. (The set clearly includes S_E since $S_E(w) = 1$, and is exactly S_E since S_E is maximal.) The answer returned by C-MBR is correct, due to Condition 12 and Theorem 8. Therefore the algorithm correctly responds "Yes". □

The following lemmas identify cases in which the required conditions hold.

Lemma 15. *If the set D of defaults consists of*

(i) *positive defaults, or*

(ii) *simple defaults with $\leqslant r$ negative literals, or*

(iii) *up to* $\log n$ *default rules,*

then there is a small basis which satisfies Condition 13.

Proof. For (i), since D is a set of positive defaults, every prime implicant of $S \subseteq D$ is a monotone term, and $b = 1^n$ satisfies the condition. That is, the Condition 13 holds for every monotone basis that contains 1^n. For (ii), every subset $S \subseteq D$ is a conjunction of literals, and S has a single prime implicant, itself. Thus, every basis which includes an assignment in which all the negative literals of S assigned 0, and all other literals assigned 1, satisfies Condition 13. In particular, if D contains up to r negative literals, so does every subset of it S, and the basis B_{H_r} suffices. For (iii), every subset $S \subseteq D$ has a CNF expression with at most $\log n$ clauses, and therefore also a DNF expression in which every term has at most $\log n$ literals. Therefore, if u satisfies S, u also satisfies a prime implicant t with no more than $\log n$ literals. Thus, if the monotone basis B contains an $(n, \log n)$ universal set, Condition 13 holds, since t must satisfy at least one of the elements there. □

The following definition captures the cases for which model-based default reasoning is correct:

Definition 16. The following cases describe simultaneous restrictions on classes $(D, \mathcal{Q}, \Gamma)$ that guarantee efficient solution for the default reasoning problem. We denote these cases by $\mathcal{TD}$ (for tractable defaults).

(i) D: a set of positive defaults; $\mathcal{Q}$: the class of k-quasi-Horn queries. The corresponding model-based representation: $\Gamma = \Gamma_W^{B_{\mathrm{H}_k}}$.

(ii) D: a set of simple defaults with up to r negative literals; $\mathcal{Q}$: the class of k-quasi-Horn queries. The corresponding model-based representation: $\Gamma = \Gamma_W^{B_{\mathrm{H}_{k+r}}}$.

(iii) D: a set of at most $\log n$ default rules (disjunctions); $\mathcal{Q}$: the class of $\log n$-CNF queries. The corresponding model-based representation: $\Gamma = \Gamma_W^{B_{2\log n\text{-CNF}}}$.

Lemma 17. *Condition* 12 *and Condition* 13 *are satisfied for all* $(D, \mathcal{Q}, \Gamma)$ *in* $\mathcal{TD}$.

Proof. Condition 12 holds as a direct consequence of Theorem 9.

Condition 13 holds as a direct implication of Lemma 15: For (i), since $1^n \in B_{\mathrm{H}_{k+1}}$. For (ii), since $B_{\mathrm{H}_r} \subset B_{\mathrm{H}_{k+r}}$, and for (iii), since $B_{2\log n\text{-CNF}}$ contains a $(n, \log n)$-universal set. □

Using Theorem 14 and Lemma 17 we get:

Corollary 18. *The algorithm CD-MBR solves the credulous default reasoning task CDEF(D, W, q) correctly, for all $q \in \mathcal{Q}$ and for all $(D, \mathcal{Q}, \Gamma)$ in $\mathcal{TD}$.*

Algorithm SD-**MBR**(Γ, D, q):

```
If for all models z ∈ Γ q(z) = 0, return "No".
Do for all models z ∈ Γ such that q(z) = 1
    Let S = {d ∈ D | d(z) = 1}
    S_flag = max                    /* Consider S as a potential maximal subset */
    Do for all d ∈ D \ S
        if there exists y ∈ Γ such that W(y) = 1, S(y) = 1, and d(y) = 1
        then S_flag = no-max
    EndDo
    If S_flag = max and C-MBR(Γ, S, q) answers "No", return "No".
EndDo
Return "Yes".                                   /* All extensions are good */
```

Fig. 4. SD-MBR: default reasoning with a model-based representation.

5.2. Skeptical default reasoning

The only difference between the credulous and the skeptical reasoning tasks is that in the latter we respond affirmatively only if the query holds in *all* extensions rather than just in one. A legitimate extension is a maximal subset of the set D of default rules. For credulous default reasoning it was sufficient to guarantee that all the maximal sets are considered as candidates, and it was not important that some non-maximal sets were considered as well (since they could not affect the response of the algorithm). For skeptical reasoning, we need to identify the maximal sets, since a consistent subset S which is not maximal might satisfy $W \wedge S \not\models q$ while every maximal subset S_E which includes S satisfies $W \wedge S_E \models q$. Therefore, the main stage of the algorithm SD-MBR, presented in Fig. 4, tests for maximal subsets.

Let $\Gamma = \Gamma_W$ be a model-based representation of W. The algorithm SD-MBR receives Γ, D and a query q as input. It starts by enumerating all the models in Γ. When it finds a model z in which the query holds (i.e., $q(z) = 1$), it sets S to be the set of all the rules in D that this model satisfies. The algorithm then tests whether S is indeed a maximal consistent subset, by checking whether there is any superset of S that is consistent with W. (In the correctness proof of the algorithm we show that it is sufficient to test this condition using elements of Γ.) If S is not maximal then it is ignored and the algorithm goes on to the next assignment in Γ. Otherwise, the algorithm tests whether $W \wedge S \models q$ by calling the procedure C-MBR to decide whether $W \models_S q$. If the answer is "No" the algorithm returns "No", and otherwise it continues to look for another maximal set S. If all the models in Γ, and the corresponding subsets of D, have been tested and no bad extension has been found then the algorithm says "Yes".

Theorem 19. *Given Γ_W^B, the algorithm SD-MBR solves the skeptical default reasoning task SDEF(D, W, q) correctly whenever Condition* 12 *and Condition* 13 *hold.*

Proof. The proof is similar to that of Theorem 14. We need to prove:

(i) if the algorithm returns "Yes" then all extensions contain q, and

(ii) if all extensions contain q, then the algorithm returns "Yes".

Consider first case (i). The proof of Theorem 14 shows that every extension S is indeed considered, for some $z \in \Gamma$. Since an extension S is maximal, it passes the inner loop test in *SD*-MBR. That is, no superset of S which is consistent with W is found. Therefore all extensions are passed to the subroutine for C-MBR, which gives a correct answer. Thus, if the algorithm returns "Yes" then all extensions contain q.

For (ii) we have that all extensions contain q. Assume first that there exists $z \in \Gamma_W^B$ such that $q(z) = 1$, and therefore the algorithm passes the first step and goes into the inner loop. We argue that all the subsets S identified as extensions (i.e., flagged *max* after the inner loop) indeed correspond to extensions. Assume that S is not maximal and that for some d, $S^* = S \wedge d$ is consistent with W. Consider the iteration in the inner loop in which d is the candidate added to S. Since S^* is consistent with W we know that there is some $u \in W$ such that $S^*(u) = 1$. But, Condition 13 holds for S^* as well, and therefore there is a $y \in \Gamma$ which satisfies S^*. Thus, the algorithm will detect this fact, set $S_{flag} = no\text{-}max$, and ignore S as required. As in case (i) we know that the answers of C-MBR are correct and thus the algorithm says "Yes" as required.

We now prove that, as assumed above, there exists $z \in \Gamma_W^B$ such that $q(z) = 1$. Notice that since all extensions contain q, for all extensions, and in particular for some extension E, there exists $x \in W \wedge S_E \wedge q$. As above this implies that there is also a $y \in \Gamma_W^B$ satisfying this, and thus the assumption holds. □

Using Theorem 19 and Lemma 17 we get:

Corollary 20. *The algorithm SD-MBR solves the skeptical default reasoning task* $SDEF(D, W, q)$ *correctly, for all* $q \in \mathcal{Q}$ *and for all* $(D, \mathcal{Q}, \Gamma)$ *in* $\mathcal{TD}$.

5.3. Application: Diagnosis using models

One of the useful applications of default logic is for the problem of circuit diagnosis [21]. Consider for example the circuit $d \leftarrow a \wedge b$; $e \leftarrow d \vee c$, composed of one and gate and one or gate. In order to diagnose possible problems in the circuit we add, for every gate, a new variable denoting that it is operating normally. In our example N_1 will correspond to the and gate and N_2 will correspond to the or gate. Using these new variables, the functionality of the circuit can be described by

$$W = (N_1 ab \to d)(N_1 d \to a)(N_1 d \to b)(N_2 c \to e)(N_2 d \to e)(N_2 e \to c \vee d).$$

Under normal conditions we want to assume that all the gates are operating normally. This is captured in our example by the set of positive simple default rules $D = \{\frac{:N_1}{N_1}, \frac{:N_2}{N_2}\}$, and in general by having one rule for each gate's "normality" variable.

In the absence of any further evidence, considering the default problem (W, D) reveals that there is only one extension, D. This should be interpreted as stating that all the gates are operating normally.

Algorithm *Diag*-MBR(Γ, O, D, q):

Let $\Gamma_O = \{z \in \Gamma \mid O(z) = 1\}$.
Call *SD*-MBR(Γ_O, D, q) and answer in the same way.

Fig. 5. *Diag*-MBR: model-based diagnostic reasoning.

Suppose, however, that we observe that $c = 1$ and $e = 0$. In such a case the circuit can be described as $W' = W \wedge c\bar{e}$, and we need to consider the default reasoning problem (W', D). It is easy to see that in this case there is only one extension which includes N_1 but does not include N_2. This should be interpreted as stating that a minimal explanation for the fault is that gate number 2 is faulty.

Of course, it is not always the case that observations exactly determine the fault in the circuit. Suppose, for example, that we observe $a = 1$, $b = 1$, and $e = 0$. In this case $W' = W \wedge ab\bar{e}$, and (W', D) has two extensions. In one extension we have N_1 but not N_2 and in the other we have N_2 but not N_1. This identifies the possible scenarios that could have caused the fault in the circuit; either the and gate or the or gate are not functioning correctly. If we want to know whether the circuit implies $d \rightarrow a$, it makes sense to use the skeptical default reasoning paradigm with (W', D) and the query $d \rightarrow a$. This has the following interpretation: "can $d \rightarrow a$ be deduced without knowing the specifics of the fault?" It is easy to see that the answer is "No" (due to the case $N_1 = 0$).

We now show how to apply our positive results for default reasoning to the problem of diagnosis. Observe that in the problem of diagnosis our knowledge about the world varies with the observations. Therefore the observations serve as a kind of context information. On top of that we have to support the default rules that capture the assumptions on the normality of the gates. This suggests the algorithm *Diag*-MBR, described in Fig. 5. The algorithm first uses filtering as in *C*-MBR, where the observations O serve as the context information. Then it uses the set of filtered models as the knowledge base for the algorithm *SD*-MBR.

As the following theorem shows, for the strategy to succeed it is sufficient that the expression $(O \rightarrow (S \rightarrow q))$ is supported as a query by the characteristic models.

Theorem 21. *The algorithm Diag-MBR solves the diagnosis task correctly as long as O includes at most r observations, q is a k-quasi-Horn expression, and* $\Gamma = \Gamma_W^{B_{\mathrm{H}_{k+r}}}$.

Proof. Given the set of observation O, the requirement is to solve the default task $SDEF(D, (W \wedge O), q)$. As in Theorem 19 we need to show that all extensions are considered by the algorithm, and that the subroutine for reasoning within context is used properly.

Let S_E be an extension of $W \wedge O$. Namely, there is a model $u \in W \wedge O \wedge S_E$. Since S_E is a subset of D it includes only positive literals. Also, by the conditions of the theorem, O has at most r literals (and therefore at most r negative literals). This implies that $B_{\mathrm{H}_{k+r}}$ includes an assignment b which agrees with both O and S_E on all literals. Thus, there is an assignment $z \in \Gamma$, $z \leqslant_b u$ such that z agrees with b and u on all the literals in S_E and O. The assignment z will survive the filtering stage of the algorithm, and the extension S_E will be properly identified. (Note that as before, if the algorithm claims that an extension is maximal then it is maximal since otherwise it would be detected due to the existence of z.)

When an extension S is identified by the algorithm *SD*-MBR, it uses a subroutine call to C-MBR to test whether the question q follows given the context S. In our case, the question is whether $(W \wedge O) \models_S q$ which is equivalent to $W \wedge O \wedge S \models q$. Using Theorem 8 we get that this is answered correctly as long as B is a basis for $\overline{O} \vee \overline{S} \vee q$. Since S is monotone, and O has at most r literals this holds in our case. □

5.4. Default reasoning and relevance

The use of defaults in reasoning is motivated by the desire to allow for efficient decision-making in the presence of incomplete information. The goal is to exploit the information given with respect to "typical" cases, namely, the default rules, when it is relevant to the current situation. Moreover, this process of identifying the relevant cases should, intuitively, contribute somehow to reducing the complexity of reasoning.

However, it turns out that when using default logic as the default reasoning framework, this goal is missed. While each rule in the representation constitutes a local assumption, a set of default rules tries to capture all plausible scenarios, and the default reasoning task is defined in relation to all these scenarios. In fact, the computational difficulties in default logic can be traced to this search for possible scenarios.

As we have shown, the relation of defaults to relevance can be viewed through the notion of context. Namely, default reasoning can be viewed as a generalization of reasoning within context where the agent has to find plausible context information through the use of default rules. We have also shown that model-based representations can, in some cases, capture all possible extensions in an accessible form. In these cases, one can consider one context at a time and use the filtering technique in order to focus on the relevant information in each of these contexts. This is shown to yield efficient default reasoning algorithms.

6. Relevance to the task

The view we take on common-sense reasoning is that the agent has to function in a very complex world that may be hard to represent exactly. Luckily, the agent need not be omniscient, but rather has to perform well on a fairly wide, but restricted, set of tasks. Given that we relax the requirements, the question is whether the agent needs a complete description of the world in order to reason correctly with respect to it, or perhaps, can perform as well using only partial information that is relevant to the task.

One way to study this issue is via our notion of relevance—a way to reduce the computational cost of reasoning. While before we used relevance or context information to prune the knowledge representation for a particular situation, it is also possible to prune a knowledge representation relative to the general task the agent is supposed to perform.

Consider the deduction problem, and suppose that our agent were to wander in a world in which all queries are restricted in some form, or belong to some language $\mathcal{Q}$. This means that the agent needs to answer correctly only queries in $\mathcal{Q}$, and may (potentially) be wrong on queries not in $\mathcal{Q}$, as it is not going to be queried on those anyway.

In this case, it is known that an incomplete description of the world is sufficient to support correct deduction. This can be formalized using the notion of a *least upper bound representation*, introduced by Kautz and Selman [27]. Intuitively, these approximations capture all the conclusions of W which belong to $\mathcal{Q}$. In [9] it is shown that these approximations support exact deduction with respect queries in $\mathcal{Q}$.

In fact, the model-based representations that we have used earlier in the paper capture such approximations relative to the class of queries. Thus, our results implicitly use the notion of *relevance to the task* as well. Moreover, the use of model-based representations is essential to exploit the relevance to the task. The reason is that the representation should be in a form that supports efficient solutions of the reasoning problem. For example, suppose the task is answering $\log n$-CNF queries. Given a CNF representation of the least upper bound of W with respect to $\log n$-CNF it is NP-hard to reason with this representation and therefore not feasible. The existence of a compact model-based representation of this least upper bound, enables us to perform the reasoning efficiently.

7. Learning to reason

The results presented in previous sections support an incremental view of reasoning in a natural way. We have shown that a model-based representation can be used to reason correctly when some additional constraining context information is supplied. This information augments the agents' knowledge and aids in deriving conclusions relevant to this context. We call this a *top-down* solution, since it assumes that the agent has a complete knowledge base, but uses only parts of it, depending on the current context.

It is conceivable, though, for an agent to have only *some* of the models, those models that come from some specific context d. In such a case, our results show that the agent reasons correctly within this context (although not within every context). We call this a bottom-up approach. This approach supports the view that an intelligent agent can construct a representation of the world incrementally by pasting together many "narrower" views from different contexts. In each of those, the agent is guaranteed to reason correctly and, eventually, it constructs a more complete knowledge base, which supports many possible contexts, even those never experienced.

This intuitive approach can be cast in a more general framework which emphasizes the inductive nature of reasoning. In systems that learn, the world in which the agent performs its task is the same world that supplies the agent with the information when learning. This intuition is captured in the distribution free model of learning theory [30]. There, an agent first wanders around in the world, observing examples drawn from some unknown distribution D which governs the occurrences of instances in the world. Then, the agent has to perform its task, namely to classify instances. The agent is allowed to err on some set of instances as long as the measure of this set under D is small. Thus, the same arbitrary "world" that supplies the information in the learning phase is used to measure the agent's performance later. This intuition was not captured by early formulations of reasoning, where the agent has an exact formula-based description of the world (traditionally, a CNF formula), and its performance is defined in some way that does not depend on the world it functions in (e.g., by the ability to make arbitrary deductions).

In [7] a general framework *Learning to Reason* is defined, that incorporates the ideas above into the study of reasoning. In this framework the intelligent agent is given access to its favorite learning interface, and is also given a grace period in which it can interact with this interface and construct its representation[6] KB of the world W. The reasoning performance is measured only after this period, when the agent is presented with queries α from some query language, relevant to the world, and has to answer whether W implies α.

This framework allows for a formal study of yet another manifestation of our notion of relevance as a way to reduce the computational cost of reasoning. In this case we may call it *relevance to the environment*. Namely, the performance of an agent has to be measured by some criterion that depends on the world the agent functions in. Indeed, it is shown in [7] that through this interaction with the world, the agent truly gains additional reasoning power.

We briefly describe two results which emphasize how, within this framework, agents can focus on relevant information and what they gain from using it. As before, these results rely on the use of model-based representations.

A sampling approach

Suppose we have access to random examples from a certain context d in our world W (e.g., the "conference" context discussed above). This may allow us to take random samples according to the distribution D that governs the occurrences of instances in $W \wedge d$. It can be shown [7] that a sample of $m = (p/\varepsilon)\ln(1/\delta)$ random examples can answer correctly all questions of length $\leqslant p$, which are not evasive. A statement is evasive if it is not implied by the world but, in practice, it is falsified very rarely. (Formally, α is evasive if $W \wedge d \not\models \alpha$ but $\Pr_D[W \wedge d \wedge \overline{\alpha}] < \varepsilon$.)

This exemplifies the notion of relevance to the environment: the environment the agent interacts with is defined by the distribution D, and the queries the agent cares about (the non-evasive queries) are defined relative to this environment.

Notice that this is very similar to the usage of model-based representations in the framework for reasoning within context. Now, instead of having a fixed and well defined set of models, which supports exact reasoning, a random set is used and we require only probably-approximately-correct reasoning.

[6] Note that in this framework we need to distinguish between the world W and the agent's representation KB.

Theory approximation and restricted queries

The utility of knowledge approximations for capturing information relevant to the task was discussed in Section 6. We note that the use of these approximations is advantageous for other reasons as well. In [7] it is shown that while exact learning of functions is still not within reach, one can learn the model-based representations for the least upper bound approximations of functions. Thus, one can learn such approximations, in the appropriate environment, and then use it for reasoning, combining the ideas of relevance to the task and relevance to the environment.

8. Conclusions

Reasoning with models is an intuitive paradigm that has been shown to be theoretically sound. In this paper we presented more evidence for the utility of model-based representations. In particular, these representations support efficient reasoning in the presence of varying context information, as well as some restricted cases of default reasoning. We further argued that the notion of relevance can be naturally and successfully utilized when reasoning with models.

The basic computational task we considered is the problem of *reasoning within a varying context*. We modeled this situation by augmenting the knowledge we have about the world with context-specific information, and showed how to solve this task efficiently using a model-based representation. Our solution, the filtering algorithm, directly implements the idea of ignoring irrelevant information.

In default reasoning, an agent may have many (possibly conflicting) default rules, acquired in different contexts. As we have shown, the relation of defaults to relevance can be viewed through the notion of context. Namely, default reasoning can be viewed as a generalization of reasoning within context where the agent has to find plausible context information through the use of default rules. Furthermore, under certain restrictions, model-based representations capture all possible scenarios in an accessible form and this can be exploited to yield efficient default reasoning algorithms.

The significance of these results is that they are achieved as natural extensions of deductive reasoning, and hold in cases in which the traditional formula-based representation does not support efficient reasoning.

Moreover, we have shown that these results support an incremental view of reasoning in a natural way, and discussed the Learning to Reason framework and the notion of relevance as manifested in it. In particular, within this framework it has been shown that the model-based representations discussed here can be learned efficiently. This can be combined with context specific default rules that are acquired via rote learning or other learning processes [24] to work in a plausible way.

These results can be viewed as providing some theoretical support for the usefulness of case-based style reasoning, where a set of "typical cases" is used as a knowledge representation. More work is needed in order to gain a deeper understanding of these issues.

We believe that an effective use of the relevant information is an important part of any efficient solution of the reasoning task. In this respect, model-based representations serve as a good example, since they allow to exploit various aspects of relevance in a natural way.

Acknowledgments

We thank the anonymous reviewers for comments that helped to improve the presentation of this paper.

References

[1] N. Alon, J. Bruck, J. Naor, M. Naor and R. Roth, Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs, *IEEE Trans. Inform. Theory* 38 (1992) 509–516.
[2] N.H. Bshouty, Exact learning via the monotone theory, *Inform. and Comput.* 123 (1995) 146–153.
[3] P.N. Johnson-Laird, *Mental Models* (Harvard University Press, Cambridge, MA, 1983).
[4] P.N. Johnson-Laird and R.M.J. Byrne, *Deduction* (Lawrence Erlbaum, London, 1991).
[5] H. Kautz, M. Kearns and B. Selman, Horn approximations of empirical data, *Artificial Intelligence* 74 (1995) 129–145.
[6] H. Kautz and B. Selman, Hard problems for simple default logics, *Artificial Intelligence* 49 (1991) 243–279.
[7] R. Khardon and D. Roth, Learning to reason, in: *Proceedings AAAI-94*, Seattle, WA (1994) 682–687.
[8] R. Khardon and D. Roth, Learning to reason with a restricted view, in: *Workshop on Computational Learning Theory* (1995) 301–310.
[9] R. Khardon and D. Roth, Reasoning with models, *Artificial Intelligence* 87 (1996) 187–213.
[10] J. Kolodner, *Case-Based Reasoning* (Morgan Kaufmann, Los Altos, CA, 1993).
[11] S.M. Kosslyn, *Image and Mind* (Harvard University Press, Cambridge, MA, 1983).
[12] H.J. Levesque, Making believers out of computers, *Artificial Intelligence* 30 (1986) 81–108.
[13] H.J Levesque, Is reasoning too hard?, in: *Proceedings 3rd NEC Research Symposium* (1992).
[14] J. McCarthy and P.J. Hayes, Some philosophical problems from the standpoint of artificial intelligence, in: B. Meltzer and D. Michie, eds., *Machine Intelligence*, Vol. 4 (Edinburgh University Press, Edinburgh, 1969).
[15] M. Minsky, A framework for representing knowledge, in: P.H. Winston, ed., *The Psychology of Computer Vision* (McGraw-Hill, New York, 1975); also in: R.J. Brachman and H.J. Levesque, *Readings in Knowledge Representation* (Morgan Kaufmann, Los Altos, CA, 1985).
[16] J. Naor and M. Naor, Small-bias probability spaces: Efficient constructions and applications, *SIAM J. Comput.* 22 (1993) 838–856.
[17] C.H. Papadimitriou, On selecting a satisfying truth assignment, in: *Proceedings 32nd Annual IEEE Symposium on Foundations of Computer Science* (1991).
[18] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann, Los Altos, CA, 1988).
[19] R. Reiter, A logic for default reasoning, *Artificial Intelligence* 13 (1980) 81–132.
[20] R. Reiter, Nonmonotonic reasoning, in: *Annual Reviews of Computer Science* (Annual Reviews Inc., Palo Alto, CA, 1987) 147–188.
[21] R. Reiter, A theory of diagnosis from first principles, *Artificial Intelligence* 32 (1987) 57–95.
[22] D. Roth, Learning to reason: the non-monotonic case, in: *Proceedings IJCAI-95*, Montreal, Que. (1995) 1178–1184.
[23] D. Roth, On the hardness of approximate reasoning, *Artificial Intelligence* 82 (1996) 273–302.
[24] D. Schuurmans and R. Greiner, Learning default concepts, in: *Proceedings 10th Canadian Conference on Artificial Intelligence (CSCSI-94)*, Banff, Alta. (1994).

[25] B. Selman, Tractable default reasoning, Ph.D. Thesis, Department of Computer Science, University of Toronto, Ont. (1990).
[26] B. Selman and H. Kautz, Model-preference default theories, *Artificial Intelligence* 45 (1990) 287–322.
[27] B. Selman and H. Kautz, Knowledge compilation and theory approximation, *J. ACM* 43 (1996) 193–224.
[28] B. Selman and H.J. Levesque, Abductive and default reasoning: a computational core, in: *Proceedings AAAI-90*, Boston, MA (1990) 343–348.
[29] D.S. Touretzky, J.F. Horty and R.H Thomason, A clash of intuitions: the current state of nonmonotonic multiple inheritance systems, in: *Proceedings IJCAI-87*, Milan, Italy (Morgan Kaufmann, Los Altos, CA, 1987).
[30] L.G. Valiant, A theory of the learnable, *Comm. ACM* 27 (1984) 1134–1142.
[31] L.G. Valiant, Rationality, in: *Workshop on Computational Learning Theory* (1995) 3–14.