Reasoning with Models

Roni Khardon* Dan Roth†

Aiken Computation Laboratory,
Harvard University,
Cambridge, MA 02138.
{roni,danr}@das.harvard.edu

Abstract

We develop a model-based approach to reasoning, in which the knowledge base is represented as a set of models (satisfying assignments) rather then a logical formula, and the set of queries is restricted. We show that for every propositional knowledge base (KB) there exists a set of *characteristic models* with the property that a query is true in KB if and only if it is satisfied by the models in this set. We fully characterize a set of theories for which the model-based representation is compact and provides efficient reasoning. These include some cases where the formula-based representation does not support efficient reasoning. In addition, we consider the model-based approach to *abductive reasoning* and show that for any propositional KB, reasoning with its model-based representation yields an abductive explanation in time that is polynomial in its size.

Introduction

A widely accepted framework for reasoning in intelligent systems is the knowledge-based system approach (McCarthy 1958). Knowledge, in some *representation language* is stored in a Knowledge Base (KB) that is combined with a reasoning mechanism. Reasoning is abstracted as a deduction task of determining whether a sentence α, assumed to capture the situation at hand, is implied from KB (denoted KB $\models \alpha$). However, computational considerations render this logical-based representation, as well as many other forms of reasoning (Selman 1990; Roth 1993), as not adequate for common-sense reasoning (Levesque 1986; Shastri 1993).

In this work we embark on the development of a model-based approach to common sense reasoning. 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 to reasoning have been cognitive psychologists (Johnson-Laird 1983; Johnson-Laird & Byrne 1991; Kosslyn 1983). In the AI community this approach can be seen, in a very general sense, as a derivative of Levesque's notion of "vivid" reasoning, and is very related to the approach developed in (Kautz, Kearns, & Selman 1993).

The problem KB $\models \alpha$ can be approached using the following model-based strategy:

Test Set: A set Γ of assignments.

Test: If there is an element $x \in \Gamma$ which satisfies KB, but does not satisfy α, deduce that KB $\not\models \alpha$; Otherwise, KB $\models \alpha$.

Clearly, (since KB $\models \alpha$ iff every model of KB is also a model of α) this approach solves the inference problem if Γ is the set of *all* models of KB. A model-based approach 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, under some criterion.

We define a set of models, the *characteristic models* of the knowledge base, with the property that performing the model-theory test on them suffices to deduce that KB $\models \alpha$, *for a restricted set of queries.* We prove that for a fairly wide class of representations, this set is sufficiently small, and thus the model-based approach is feasible. The notion of *restricted queries* is inherent to our approach. Since we are interested in formalizing common-sense reasoning, we take the view that a reasoner need not answer efficiently *all* possible queries.

For a wide class of queries we show that exact reasoning can be done efficiently, even when the reasoner keeps in KB only an "approximate" representation (as a set of characteristic models) of the "world". We show that the theory developed here generalizes the model-based approach to reasoning with Horn theories, studied in (Kautz, Kearns, & Selman 1993), and captures even the notion of reasoning with approximate theories (Selman & Kautz 1991). In particular, our results characterize the Horn theories for which the approach in (Kautz, Kearns, & Selman 1993) is useful, and explain the phenomena observed there, regarding the relative sizes of the logical formula representation and model-based representation of KB. We also give

*Research supported by grant DAAL03-92-G-0164 (Center for Intelligent Control Systems).

†Research supported by NSF grant CCR-92-00884 and by DARPA AFOSR-F4962-92-J-0466.

other examples of expressive families of propositional theories, for which our approach is useful.

In addition, we consider the problem of performing *abduction* using a model-based approach and show that for any propositional knowledge base, using a model-based representation yields an abductive explanation in time that is polynomial in the size of the model-based representation. Some of our technical results make use of a new characterization of Boolean functions, called the *Monotone Theory*, introduced recently by Bshouty (Bshouty 1993). Due to the limited space, some of the proofs are omitted. These can be found in the full version of the paper (Khardon & Roth 1994b).

Summary of Results

We now briefly describe the main applications of the model-based approach developed in this paper.

We consider two types of queries with which reasoning is efficient. Queries are called *relevant* if they belong to the propositional language that represents the "world". Queries are called *common* if they belong to some set $\mathcal{L}_E$ of *efficient* propositional languages (see Definition 5). These include for example Horn queries, and $\log n$CNF queries.

Our results can be grouped into 3 categories that can be informally described as follows:
(1) Every function with a small DNF representation and either a small CNF representation or a CNF representation (of any size) in $\mathcal{L}_E$ has a small set of characteristic models.

For these functions, model-based deduction is correct and efficient for relevant and for common queries.
(2) The set $\Gamma^{\mathcal{G}}$, of characteristic models with respect to a propositional language $\mathcal{G}$, describes the least upper bound of f with respect to $\mathcal{G}$.

Model-based deduction, using $\Gamma^{\mathcal{G}}$, is correct and efficient for common queries.
(3) For the functions defined in (1), efficient and correct model-based abduction can be performed.

We note that our algorithms do not solve NP-complete problems. Most hardness results for reasoning assume that KB is given as a CNF formula. The fact that we can perform reasoning efficiently relies on the fact that we change the knowledge representation into a more accessible form (another knowledge representation which enables reasoning, yet for some reason is considered less interesting, is DNF).

Monotone Theory

In this section we introduce the notations, definitions and results of the Monotone Theory of Boolean functions (Bshouty 1993).

We consider a Boolean function $f : \{0,1\}^n \rightarrow \{0,1\}$. An *assignment* simply means an element of $\{0,1\}^n$. A *model* of f is a *satisfying assignment* of f i.e., x such that $f(x) = 1$. Throughout the paper, when no confusion can arise, we identify f with the set of its models, namely $f^{-1}(1)$. That is, $f \models g$ if and only if $f \subseteq g$. Assignments in $\{0,1\}^n$ are denoted by x, y, z, and x_i denotes the ith coordinate of $x \in \{0,1\}^n$.

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

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 and we have that $1 <_{b_i} 0$.
The *monotone extension of* $z \in \{0,1\}^n$ with respect to b is:

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

The *monotone extension of* f with respect to b is:

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

The set of *minimal assignments of* f with respect to b is:

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

The following claims list some properties of $\mathcal{M}_b$.

Claim 1 *Let* $f, g : \{0,1\}^n \rightarrow \{0,1\}$ *be Boolean functions. The operator* $\mathcal{M}_b$ *satisfies the following properties:*
(i) If $f \subseteq g$ *then* $\mathcal{M}_b(f) \subseteq \mathcal{M}_b(g)$.
(ii) $\mathcal{M}_b(f \wedge g) \subseteq \mathcal{M}_b(f) \wedge \mathcal{M}_b(g)$.
(iii) $\mathcal{M}_b(f \vee g) = \mathcal{M}_b(f) \vee \mathcal{M}_b(g)$.

Claim 2 *Let* $z \in f$. *Then, for every* $b \in \{0,1\}^n$, *there exists* $u \in min_b(f)$ *such that* $\mathcal{M}_b(z) \subseteq \mathcal{M}_b(u)$.

From Claims 1 and 2 we get a characterization of the monotone extension of f:

Claim 3 *The monotone extension of* f *with respect to* b *is:*

$$\mathcal{M}_b(f) = \bigvee_{z \in f} \mathcal{M}_b(z) = \bigvee_{z \in min_b(f)} \mathcal{M}_b(z).$$

Clearly, for every assignment $b \in \{0,1\}^n$, $f \subseteq \mathcal{M}_b(f)$. Moreover, if $b \notin f$, then $b \notin \mathcal{M}_b(f)$ (since b is the smallest assignment with respect to the order $\leq_b$). Therefore:

$$f = \bigwedge_{b \in \{0,1\}^n} \mathcal{M}_b(f) = \bigwedge_{b \notin f} \mathcal{M}_b(f).$$

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* $\mathcal{F}$.

Using this definition, the representation

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

yields the following necessary and sufficient condition describing when $x \in \{0,1\}^n$ is positive for f:

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

The following claim bounds the size of the basis of a function f:

Claim 4 *Let $f = C_1 \wedge C_2 \wedge \cdots \wedge C_k$ be a CNF representation for f and let B be a set of assignments in $\{0,1\}^n$. If every clause C_i is falsified by some $b \in B$ then B is a basis for f. In particular, f has a basis of size $\leq k$.*

The set of *floor* assignments of an assignment x, with respect to the order relation b, denoted $\lfloor x \rfloor_b$, is the set of all elements $z <_b x$ such that there does not exist y for which $z <_b y <_b z$ (i.e., z is strictly smaller than x relative to b and is different from x in exactly one bit).

The set of *local minimal assignments of f* with respect to b is:

$$\min_b^*(f) = \{x \mid x \in f, \text{ and } \forall y \in \lfloor x \rfloor_b, \;\; y \notin f\}.$$

Clearly we have that $\min_b(f) \subseteq \min_b^*(f)$ and therefore the following lemma bounds the size of $\min_b(f)$.

Claim 5 *Let $f = D_1 \vee D_2 \vee \cdots \vee D_k$ be a DNF representation for f. Then for every $b \in \{0,1\}^n$, $|min_b^*(f)| \leq k$.*

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: $\{0000, 0001, 0010, 1101\}$. Using Claim 4 we get that the set[1] $B = \{0000, 1101\}$ is a basis for f.

The sets of minimal assignments with respect to this basis are: $\min_{0000}(f) = \{1000, 0100, 0011\}$ and $\min_{1101}(f) = \{1100, 1111, 1001, 0101\}$. These can be easily found by drawing the corresponding lattices and checking which of the satisfying assignments of f are minimal. It is also easy to check that f can be represented as in equation (1) using the minimal elements identified.

Deduction with Models

We consider the deduction problem KB$\models \alpha$. KB is the knowledge base, which is taken to be a propositional expression (i.e., some Boolean function[2]), and α is also a propositional expression. The assertion KB$\models \alpha$ means that every model $x \in \{0,1\}^n$ which satisfies KB, must also satisfy α.

[1] 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 $x_3 = x_4 = 1$).

[2] We use interchangeably the terms propositional expression and Boolean function. Similarly, a family of Boolean functions is used interchangeably with a propositional language. A family of Boolean functions is uniquely characterized as a set of all functions with a given basis.

In this section we define a special collection Γ of *characteristic models* of KB and show that performing the model-based test on Γ yields correct deduction. We fully characterize Γ in terms of the Boolean function KB and the query α.

Exact Deduction

Definition 3 *Let $\mathcal{F}$ be a class of functions, and let B be a basis for $\mathcal{F}$. For a knowledge base $KB \in \mathcal{F}$ we define the set $\Gamma = \Gamma^B_{KB}$ of* characteristic models *to be the set of all minimal assignments of KB with respect to the basis B. Formally,*

$$\Gamma^B_{KB} = \cup_{b \in B}\{z \in min_b(KB)\}.$$

Before showing that Γ has the required properties we discuss the size of the model-based representation. The following result is immediate from Claim 5.

Lemma 1 *Let B be a basis for the knowledge base KB, and denote by $|DNF(KB)|$ the size of its DNF representation. Then, the size of a model-based representation of a knowledge base KB is*

$$|\Gamma^B_{KB}| \leq \sum_{b \in B} |min_b(KB)| \leq |B| \cdot |DNF(KB)|.$$

We note that this bound is not tight. There are cases where this bound is exponential and Γ is small.

Theorem 1 *Let $KB, \alpha \in \mathcal{F}$ and let B be a basis for $\mathcal{F}$. Then $KB \models \alpha$ iff for every $u \in \Gamma^B_{KB}$, $\alpha(u) = 1$.*

Proof: Clearly, $\Gamma = \Gamma^B_{KB} \subseteq$ KB and therefore, if there exists $z \in \Gamma$ such that $\alpha(z) = 0$ then KB $\not\models \alpha$. For the other direction assume that for all $u \in \Gamma$, $\alpha(u) = 1$. We will show that if $y \in$KB, then $\alpha(y) = 1$. From Corollary 1, since B is a basis for α and for all $u \in \Gamma$ $\alpha(u) = 1$, we have that

$$\forall u \in \Gamma, \;\; \forall b \in B, \;\; \exists v_{u,b} \in \min_b(\alpha) \text{ s.t. } u \geq_b v_{u,b}. \quad (2)$$

Consider now a model $y \in$KB. Again, Corollary 1 implies that

$$\forall b \in B, \;\; \exists z \in \min_b(KB) \text{ s.t. } y \geq_b z. \quad (3)$$

By the assumption, since $\min_b(KB) \subseteq \Gamma$, all the elements z identified in Equation 3 satisfy α and therefore, as in Equation 2 we have that

$$\forall z \in \min_b(KB), \;\; \exists v_{z,b} \in \min_b(\alpha) \text{ s.t. } z \geq_b v_{z,b}. \quad (4)$$

Substituting Equation 4 into Equation 3 gives the required condition on $y \in$KB:

$$\forall b \in B, \;\; \exists v_{(z),b} \in \min_b(\alpha) \text{ s.t. } y \geq_b v_{(z),b}$$

which implies, by Corollary 1, that $\alpha(y) = 1$. ∎

The above theorem assumed that KB and α could be described by the same basis B. This requirement is somewhat relaxed in the following theorem.

Theorem 2 *Let KB be a propositional theory with basis B and let α be a query with basis B'. Then $KB \models \alpha$ if and only if for every $u \in \Gamma^{B \cup B'}_{KB}$, $\alpha(u) = 1$.*

Proof: It is clear, from Eq. 1 and the fact that for all g and b, $g \subseteq \mathcal{M}_b(g)$, that $B \cup B'$ is a basis both for KB and α. Therefore, Theorem 1 implies the result. ■

Example: (continued) The set Γ built for our basis is: $\Gamma = \{1000, 0100, 0011, 1100, 1111, 1001, 0101\}$. Note that it includes only 7 out of the 12 satisfying assignments of f. Since model-based deduction does not make mistakes on queries implied by f we concentrate in our examples on queries not implied by f.

To exemplify Theorem 1 consider the query $\alpha_1 = \overline{x_2}\,\overline{x_3} \rightarrow x_4$. This is equivalent to $x_2 \vee x_3 \vee x_4$ which is falsified by 0000 so our B is a basis for α. Reasoning with Γ will find the counterexample 1000 and will therefore conclude $f \not\models \alpha_1$.

The query $\alpha_2 = x_1x_3 \rightarrow x_2x_4$ is equivalent to $\overline{x_1} \vee x_2 \vee \overline{x_3} \vee x_4$ which is not falsified by our basis therefore model-based deduction might be wrong. Indeed reasoning with Γ will not find a counterexample and will conclude $f \models \alpha_2$ (it is wrong since the assignment 1010 satisfies f but not α_2).

Next, to exemplify Theorem 2 consider adding a basis element for α_2. This element is 1010. The set of additional minimal elements in Γ is $\{1010\}$, and reasoning with Γ would be correct on α_2.

Approximate Theories

We now consider the case in which the set of characteristic models of KB is constructed with respect to a basis B that is *not* a basis for the knowledge base KB. This representation coincides with the notion of a least upper bound of a theory, introduced in (Selman & Kautz 1991; Kautz & Selman 1991; 1992) in the context of knowledge compilation.

Definition 4 (Least Upper-bound) *Let $\mathcal{F}, \mathcal{G}$ be families of propositional languages. Given $f \in \mathcal{F}$ we say that $f_{lub} \in \mathcal{G}$ is a $\mathcal{G}$-least upper bound of f iff $f \subseteq f_{lub}$ and there is no $f' \in \mathcal{G}$ such that $f \subset f' \subset f_{lub}$.*

These bounds are called *$\mathcal{G}$-approximations* of the original theory f. The next theorem characterizes the $\mathcal{G}$-LUB of a function and shows that it is unique.

Theorem 3 *Let f be any propositional theory and $\mathcal{G}$ a class of all propositional theories with basis B. Then*

$$f_{lub} = \bigwedge_{b \in B} \mathcal{M}_b(f).$$

Proof: Define $g = \bigwedge_{b \in B} \mathcal{M}_b(f)$. We need to prove that (1) $g \subseteq f$, (2) $g \in \mathcal{G}$ and (3) there is no $f' \in \mathcal{G}$ such that $f \subset f' \subset f_{lub}$. (1) is immediate from Claim 1. To prove (2) we need to show that B is a basis for g. Indeed,

$$\begin{aligned}\bigwedge_{b \in B} \mathcal{M}_b(g) &= \bigwedge_{b \in B} \mathcal{M}_b(\bigwedge_{b \in B} \mathcal{M}_b(f)) \\ &\subseteq (\bigwedge_{b \in B} \mathcal{M}_b(f)) \bigwedge (\bigwedge_{b_i \neq b_j} \mathcal{M}_{b_i}\mathcal{M}_{b_j}(f)) \\ &= g \bigwedge (\bigwedge_{b_i \neq b_j} \mathcal{M}_{b_i}\mathcal{M}_{b_j}(f)) \subseteq g.\end{aligned}$$

Since in general $g \subseteq \bigwedge \mathcal{M}_b(g)$ we get that $\bigwedge_{b \in B} \mathcal{M}_b(g) = g$ and therefore $g \in \mathcal{G}$. Finally, to prove (3) assume that there exists $f' \in \mathcal{G}$ such that $f \subseteq f'$. Then,

$$g = \bigwedge_{b \in B} \mathcal{M}_b(f) \subseteq \bigwedge_{b \in B} \mathcal{M}_b(f') = f',$$

where the last equality results from the fact that $f' \in \mathcal{G}$. Therefore, $g = f_{lub}$. ■

The following theorem can be seen as a generalization of Theorem 1, in which we do not require that the basis B is the basis of KB. A weaker version of the corollary that follows, for the case in which $\mathcal{G}$ is the class of Horn theories, is discussed in (Kautz & Selman 1991; Cadoli 1993).

Theorem 4 *Let $KB \in \mathcal{F}$, $\alpha \in \mathcal{G}$ and let B be a basis for $\mathcal{G}$. Then $KB \models \alpha$ if and only if for every $u \in \Gamma^B_{KB}$, $\alpha(u) = 1$.*

Proof: We have shown in Theorem 3 that

$$KB_{lub} = \bigwedge_{b \in B} \mathcal{M}_b(KB) = \bigwedge_{b \in B} \bigvee_{z \in \min_b(KB)} \mathcal{M}_b(z).$$

By Theorem 1, since $\alpha(u) = 1$ for every $u \in \Gamma^B_{KB}$, we have that $KB_{lub} \models \alpha$ and therefore $\mathrm{KB} \models \alpha$. On the other hand, since $\Gamma^B_{KB} \subseteq \mathrm{KB}$, if for some $u \in \Gamma^B_{KB}$, $\alpha(u) = 0$, $\mathrm{KB} \not\models \alpha$. ■

Corollary 2 *Reasoning with the least upper bound (with respect to the language $\mathcal{G}$) of a theory KB is correct for all queries in $\mathcal{G}$.*

Example: (continued) The Horn basis for our example is: $B_H = \{1111, 1110, 1101, 1011, 0111\}$ (see Claim 6). The minimal elements with respect to 1101 were given before. Each of 1111,0111,1011,1110 satisfies f and therefore for each of these, $\min_b(f) = b$ and together we get that $\Gamma^{B_H}_f = \{1111, 0111, 1011, 1100, 1001, 0101, 1110\}$.

For the query $\alpha_2 = x_1\ x_3 \rightarrow x_2x_4$, which is not Horn, reasoning with $\Gamma^{B_H}_f$ will be wrong. For the Horn query $\alpha_2 = x_1\ x_3 \rightarrow x_2$, reasoning with $\Gamma^{B_H}_f$ will find the counterexample 1011 and therefore be correct.

Applications

In the previous section we developed the general theory for model-based deduction. In this section we discuss applications of this theory. In particular we apply it to the case of Horn knowledge base and show that earlier work on a model-based approach, in the narrower context of Horn knowledge bases (Kautz, Kearns, & Selman 1993) coincides with our theory.

Our basic result (Theorem 1) assumed that the knowledge base and the query share the same basis. A query with this property is called a *relevant query.*

We say that queries which are taken from some propositional family with a known basis, are *common queries.* In particular, queries are *common* if they belong to a set $\mathcal{L}_E$ of *efficient* propositional languages.

Definition 5 *The set $\mathcal{L}_E$ of* efficient *propositional languages is the set of languages for which there is a small (polynomial size) fixed basis.*

Important examples of efficient languages are: (1) Horn-CNF formulas, (2) reversed Horn-CNF formulas (CNF with clauses containing at most one *negative* literal), (3) k-quasi-Horn formulas (a generalization of Horn theories in which there are at most k positive literals in each clause), (4) k-quasi-reversed-Horn formulas and (5) $\log n$CNF formulas (CNF in which the clauses contain at most $O(\log n)$ literals). Any formula that can be represented as a CNF with clauses from any combination of the above is also in $\mathcal{L}_E$. The first four can be derived from Claim 6 and the last from Claim 7 ($weight(u)$ denotes the number of 1 bits in u).

Claim 6 *The set $B_H = \{u \in \{0,1\}^n \mid weight(u) \geq n-1\}$ is a basis for any Horn CNF function.*

Claim 7 ((Bshouty 1993)) *There is a polynomial size basis for the set of* $\log n$*CNF theories.*

In the case of common or relevant queries, reasoning involves the evaluation of a propositional formula on a polynomial number of assignments. This is a very simple and easily parallelizable procedure. Moreover, Theorem 4 shows that in order to reason with common queries, we need not use the basis of KB at all, and it is enough to represent KB by the set of characteristic models with respect to the basis of the query language (one of the languages in $\mathcal{L}_E$). Claim 6 and Claim 7 together with Lemma 1 and Theorems 1,2,3 and 4 imply the following general applications of our theory:

Theorem 5 *Any function $f : \{0,1\}^n \rightarrow \{0,1\}$ that has a polynomial sized representation in both DNF and CNF form can be described with a polynomial size set of characteristic models.*

Theorem 6 *Any $f : \{0,1\}^n \rightarrow \{0,1\}$ with (any size) CNF representation in $\mathcal{L}_E$ and a polynomial size DNF representation can be described with a polynomial size set of characteristic models.*

Theorem 7 *Let KB be a knowledge base (on n variables) that can be described with a polynomial size set Γ of characteristic models. Then, for any relevant or common query, model-based deduction using Γ, is both correct and efficient.*

Theorem 8 *Let KB be a knowledge base (on n variables) that can be described with a polynomial size DNF. Then there exists a fixed, polynomial size set of models Γ, such that for any common query, a model-based deduction using Γ, is both correct and efficient.*

Horn Theories

We consider the case of Horn formulae and show that in this case our notion of *characteristic models* coincides with the notion introduced in (Kautz, Kearns, & Selman 1993).

Furthermore, our results explain the relation between sizes of the model-based and the formulae-based representations. In (Kautz, Kearns, & Selman 1993) examples are given for large Horn theories with a small set of characteristic models and vice versa, but it was not yet understood when and why it happens. Our results imply that the set of characteristic models of a Horn theory is small if the size of a DNF description for the same theory is small. The other direction is however not true (i.e., there are Horn theories with a small set of characteristic models but an exponential size DNF). In the full version we explain this phenomena in more detail. Let $char_H(KB)$ be the set of models defined in (Kautz, Kearns, & Selman 1993).

Theorem 9 *Let KB be a Horn theory and $B_H = \{u \in \{0,1\}^n \mid weight(u) \geq n-1\}$. Then, $char_H(KB) = \Gamma_{KB}^{B_H}$.*

We note, that in (Kautz, Kearns, & Selman 1993) the deduction theorem was extended to answer any query (and not just a restricted set of queries as we do here). This extension relies on a special property of Horn formulae and does not hold as is in the general case. In the full version of the paper we explain this phenomena too.

Abduction with Models

We consider in this section the question of performing abduction using a model-based representation. In (Kautz, Kearns, & Selman 1993) it is shown that for a Horn theory KB, abduction can be done in polynomial time using characteristic models. In this section we show that if we add a few base assignments to our basis, the algorithm presented there works in the general case too.

Abduction is the task of finding a minimal explanation to some observation. Formally, the reasoner is given a knowledge base KB (the *background theory*), a set of propositional letters A, (the *assumption set*), and a query letter q. An *explanation* of q is a minimal subset $E \subseteq A$ such that

1. $KB \wedge ((\wedge_{x \in E} x) \models q)$ and
2. $KB \wedge (\wedge_{x \in E} x) \neq \Phi$.

Thus, abduction involves tests for entailment and consistency, but also a search for an explanation that passes both tests.

Theorem 10 *Let KB be a background propositional theory with a basis B, let A be an assumptions set and q be a query. Let $B_H = \{x \in \{0,1\}^n | weight(x) \geq n-1\}$. Then, using the set of characteristic models $\Gamma = \Gamma_{KB}^{B \cup B_H}$ one can find an abductive explanation of q in time polynomial in $|\Gamma|$ and $|A|$.*

Proof: We use the algorithm *Explain* suggested in (Kautz, Kearns, & Selman 1993) for the case of a Horn knowledge base and show that in order for it to work in the general case it is sufficient to add the Horn basis B_H and the characteristic models relative to this basis.

The abduction algorithm *Explain* starts by enumerating all the characteristic models. When it finds a model in which the query holds, (i.e., $q = 1$) it sets E to be the conjunction of all the variables in A that are set to 1 in that model. (This is the strongest set of assumptions that are valid in this model.)

The algorithm then performs the entailment test ((1) in the definition above) to check whether E is a valid explanation. This test is equivalent to testing the deduction KB $\models (q \vee (\vee_{x \in E} \bar{x}))$, that is a deductive inference with a Horn clause as the query. According to Theorem 2 this can be done efficiently with $\Gamma_{KB}^{B \cup B_H}$.

If the test succeeds, the assumption set is minimized in a greedy fashion by eliminating variables from E and using the entailment test again. It is clear that if the algorithm outputs a *minimal* assumption set E (in the sense that no subset of E is a valid explanation, not necessarily of smallest cardinality) then it is correct. It remains to show that if an explanation exists, the algorithm will find one. To prove this, it is sufficient to show that in such a case there exists a model $x \in \Gamma$ in which both the bit q and a superset of E are set to 1.

The existence of x is a direct consequence of including the base assignment $b = 1^n$ in the basis. This is true as relative to b we have $1 <_b 0$ for each bit. Therefore if there exists an explanation y, either it is a minimal assignment relative to b, or $\exists x \leq_b y$ and x is in Γ. ∎

Conclusions and Further Work

This paper develops a formal theory of model-based reasoning. We show that a simple model-based approach can support exact deduction and abduction even when an exponentially small portion of the model space is tested. Our approach builds on (1) the characterization of a set of models of the knowledge base that captures all the information needed to reason with (2) a restricted set of queries. We prove that for a fairly large class of propositional theories, including theories that do not allow efficient formula-based reasoning, the model-based representation is compact and provides efficient reasoning.

The restricted set of queries, which we call *relevant queries* and *common queries*, can come from a wide class of efficient propositional languages, (and include, for example, quasi-Horn theories and $\log n$CNF), or from the same propositional language that represents the "world". We argue that this is a reasonable approach to take in the effort to give a computational theory that accounts for both the speed and flexibility of common-sense reasoning.

The usefulness of the approach developed here is exemplified by the fact that it explains, generalizes and unifies many previous investigations, and in particular the fundamental works on reasoning with Horn models (Kautz, Kearns, & Selman 1993) and Horn approximations (Selman & Kautz 1991; Kautz & Selman 1991; 1992). We are currently studying extensions of this theory for first order logic formalizations, and application of the theory to planning.

This work is part of a more general framework which views *learning* as an integral part of the reasoning process. We believe that some of the difficulties in constructing an adequate computational theory to reasoning result from the fact that these two tasks are viewed as separate. In (Khardon & Roth 1994a) we discuss the issue of "learning to reason" and illustrate the importance of the model-based approach for this problem.

Acknowledgments

We wish to thank Les Valiant for helpful discussions and comments.

References

Bshouty, N. H. 1993. Exact learning via the monotone theory. In *Proceedings of the IEEE Symp. on Foundation of Computer Science*, 302–311.

Cadoli, M. 1993. Semantical and computational aspects of Horn approximations. In *Proceedings of the International Joint Conference of Artificial Intelligence*, 39–44.

Johnson-Laird, P. N., and Byrne, R. M. J. 1991. *Deduction.* Lawrence Erlbaum Associates.

Johnson-Laird, P. N. 1983. *Mental Models.* Harvard Press.

Kautz, H., and Selman, B. 1991. A general framework for knowledge compilation. In *Proceedings of the International Workshop on Processing Declarative Knowledge, Kaiserlautern, Germany.*

Kautz, H., and Selman, B. 1992. Forming concepts for fast inference. In *Proceedings of the National Conference on Artificial Intelligence*, 786–793.

Kautz, H.; Kearns, M.; and Selman, B. 1993. Reasoning with characteristic models. In *Proceedings of the National Conference on Artificial Intelligence*, 34–39.

Khardon, R., and Roth, D. 1994a. Learning to reason. In these Proceedings.

Khardon, R., and Roth, D. 1994b. Reasoning with models. Technical Report TR-1-94, Aiken Computation Lab., Harvard University.

Kosslyn, S. M. 1983. *Image and Mind.* Harvard Press.

Levesque, H. 1986. Making believers out of computers. *Artificial Intelligence* 30:81–108.

McCarthy, J. 1958. Programs with common sense. In Brachman, R., and Levesque, H., eds., *Readings in Knowledge Representation, 1985.* Morgan-Kaufmann.

Roth, D. 1993. On the hardness of approximate reasoning. In *Proceedings of the International Joint Conference of Artificial Intelligence*, 613–618.

Selman, B., and Kautz, H. 1991. Knowledge compilation using Horn approximations. In *Proceedings of the National Conference on Artificial Intelligence*, 904–909.

Selman, B. 1990. *Tractable Default Reasoning.* Ph.D. Dissertation, Department of Computer Science, University of Toronto.

Shastri, L. 1993. A computational model of tractable reasoning - taking inspiration from cognition. In *Proceedings of the International Joint Conference of Artificial Intelligence*, 202–207.