Horn Approximations of Empirical Data

Henry Kautz, Michael Kearns, and Bart Selman

AI Principles Research Dept., AT&T Bell Laboratories, Murray Hill, NJ 07974
email: {kautz, mkearns, selman}@research.att.com

Abstract

Formal AI systems traditionally represent knowledge using logical formulas. Sometimes, however, a model-based representation is more compact and enables faster reasoning than the corresponding formula-based representation. The central idea behind our work is to represent a large set of models by a subset of *characteristic* models. More specifically, we examine model-based representations of Horn theories, and show that there are large Horn theories that can be exactly represented by an exponentially smaller set of characteristic models.

We show that deduction based on a set of characteristic models requires only polynomial time, as it does using Horn theories. More surprisingly, *abduction* can be performed in polynomial time using a set of characteristic models, whereas abduction using Horn theories is NP-complete. Finally, we discuss algorithms for generating efficient representations of the Horn theory that best approximates a general set of models.

1 Introduction

Logical formulas are the traditional means of representing knowledge in formal AI systems [17]. The information implicit in a set of logical formulas can also be captured by expliciting recording the set of models (truth assignments) that satisfy the formulas. However, when dealing with incomplete information, the set of models is generally much too large to be represented explicitly, because a different model is required for each possible state of affairs. Logical formulas can often provide a compact representation of such incomplete information.

There has, however, been a growing dissatisfaction with the use of logical formulas in actual applications, both because of the difficulty in writing consistent theories, and the tremendous computation problems in reasoning with them. An example of the reaction against the traditional approach is the growing body of research and applications using case-based reasoning (CBR) [14]. By identifying the notion of a "case" with that of a "model", we can view the

Preprint submitted to Elsevier Science *26 October 1994*

CBR enterprise as an attempt to bypass (or reduce) the use of logical formulas by storing and directly reasoning with a set of models. [1]

This paper explores, from a complexity-theoretic standpoint, the question of how a model-based representation could be a practical alternative to a formula-based representation in the context of incomplete information. The central idea behind our work is to represent a large set of models by a subset of *characteristic* models, from which all others can be generated efficiently. More specifically, we examine model-based representations of Horn theories.

The paper begins by comparing the size of representations. We show that there are large Horn theories that can be exactly represented by exponentially smaller sets of characteristic models. The characteristic model representation, however, is not always smaller; we also provide an example where the clausal representation is exponentially smaller than the set of characteristic models. Both the characteristic model and clausal representations are strictly better than a simple enumeration of all models. [2]

Next, we consider the complexity of reasoning with sets of characteristic models. Deduction based on a set of characteristic models takes only polynomial time, as it does using Horn theories [6]. More surprisingly, *abduction* can be performed in polynomial time using a set of characteristic models, whereas abduction using Horn theories is NP-complete [23]. This result is particularly interesting because very few other tractable classes of abduction problems are known [3,8,24].

The final part of this paper examines the problem of converting a set of models into an efficient representation. This general task can be viewed as identifying meaningful, computationally-attractive structures in a set of empirical data, where each model represents a data point [5]. As such structure identification is a way of formalizing (some kinds of) scientific discovery, where a representation is judged to be good if it compactly represents the data and can be reasoned with easily. We consider the specific problem of converting a set of models into either a set of characteristic models or a set of Horn clauses. Previously, Dechter and Pearl [5] have shown that it is easy to check when an exact translation is possible, and that in that case both kinds of representations can be generated in polynomial time. Some sets of models, however, can only be *approximated* by such representations. Converting a set of models into an approximating set of characteristic models remains easy. Dechter and Pearl provided an algorithm for the special case where the theory and

[1] This is, of course, an oversimplified description of CBR; most CBR systems incorporate both a logical background theory and a set of cases.

[2] Some closely related results have been obtained in the databased community in the development of the theory of Armstrong relations and functional dependencies [2,16]. We thank Heikki Mannila for this observation.

it's approximation have nearly the same number of models (up to a constant multiple), as well as a general algorithm for generating approximate clausal representations where each Horn clause is limited to a specified length k. They noted, however, that such k-Horn approximations can be weak — and in fact, we will demonstrate a class of theories with good Horn approximations but overly general k-Horn approximations. We therefore conclude by providing a randomized algorithm that generates a Horn theory that is arbitrarily close to the best Horn approximation, in time that is polynomial in the output size and the permissible degree of error.

2 Horn Theories and Characteristic Models

We assume a standard propositional language, and use a, b, c, d, p, and q to denote propositional variables. In any context the number of variables is finite, and is usually denoted by n. A *literal* is either a propositional variable, called a positive literal, or its negation, called a negative literal. A *clause* is a disjunction of literals, and can be represented by the set of literals it contains. A clause C *subsumes* a clause C' iff all the literals in C appear in C'. A set (conjunction) of clauses is called a *clausal theory*, and is represented by Σ. The length of a clause is the number of literals it contains, and the length of a theory Σ, written $|\Sigma|$, is the sum of the lengths of its clauses. A clause is *Horn* if and only if it contains at most one positive literal; a set of such clauses is called a *Horn theory*. (Note that we are not restricting our attention to definite clauses, which contain exactly one positive literal. A Horn clause may be completely negative.)

A *model* is a complete truth assignment for the variables that appear in the theory under consideration (equivalently, a mapping from the variables to $\{0, 1\}$). For example, the fact that m assigns the variable "x" to true can be written as "m(x)=1". We sometimes write a model as a bit vector, *e.g.*, [010...], to indicate that variable a is assigned false, b is assigned true, c is assigned false, etc.

A model *satisfies* a theory if the the theory evaluates to "true" in the model. The "models of a theory Σ", denoted by *models*(Σ), is the set of models that satisfy the theory.

If m and m' are models over the same set of variables and all the variables assigned true by m are assigned true by m', then we write $m \sqsubseteq m'$. Where M is a set of models, $|M|$ is the cardinality of the set. Using the bit vector representation, the size of the representation of M is $n|M|$.

A Horn theory Σ is a Horn upper-bound of a given set of models M if and

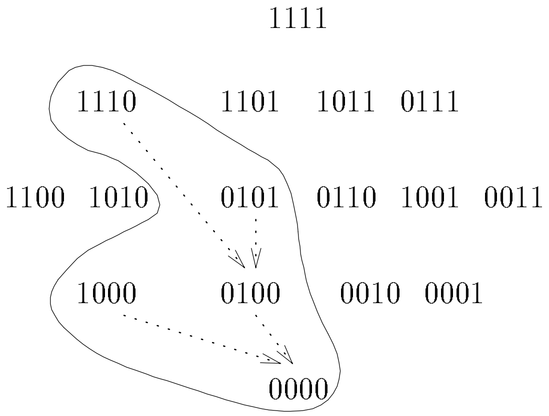

Fig. 1. The circled models are M_0', which is the closure of the example set of models M_0.

only if its models contain M:

$$M \subseteq models(\Sigma)$$

The upper-bound with the fewest models is called the *Horn approximation* of M, and corresponds to the notion of the "least upper-bound" defined in [21,22]. The Horn approximation of any set of models is unique up to logical equivalence.

It is useful for our purposes to develop an alternative but equivalent model-theoretic characterization of a Horn approximation. We begin by defining the *intersection* of a pair of models as the model that assigns "true" to just those variables that are assigned "true" by both of the pair. The *closure* of a set of models is obtained by repeatedly adding the intersection of the elements of the set to the set until no new models are generated.

Definition 1 Intersection and Closure
The intersection of models m_1 and m_2 over a set of variables is given by

$$[m_1 \cap m_2](x) \stackrel{\text{def}}{=} \begin{cases} 1 \ \textit{if} \ m_1(x) = m_2(x) = 1 \\ 0 \ \textit{otherwise} \end{cases}$$

Where M is a set of models, $closure(M)$ is the smallest set containing M that is closed under $\cap$.

To illustrate the various definitions given in this section, we will use an example set M_0 of models throughout. Let $M_0 = \{[1110], [0101], [1000]\}$. The closure of this set is given by $M_0' = M_0 \cup \{[0100], [0000]\}$. See Fig. 1.

The notion of closure is particularly relevant in the context of Horn theories, due to the following theorem.

Theorem 2 (McKinsey 1943 [18], Dechter and Pearl 1992 [5]) *A theory Σ is equivalent to a Horn theory if and only if $models(\Sigma)$ is closed under intersection.*

The original proof by McKinsey is for first-order equational theories, and in fact led to the original definition of Horn clauses [9]. A direct proof for the propositional case appears in [5].

Thus there is a direct correspondence between Horn theories and sets of models that are closed under intersection. For example, consider the closure M_0' of the models in set M_0 defined above. It is not difficult to verify that the models in the closure are exactly the models of the Horn theory $\Sigma_0 = \{\neg a \vee \neg b \vee c, \neg b \vee \neg c \vee a, \neg a \vee \neg d, b \vee \neg d, b \vee \neg c\}$.

The closure property is also useful in the characterization of Horn approximations. In fact, the closure of a set of models exactly yields the set of models of its Horn approximation:

Theorem 3 *For any set of models M with Horn approximation Σ, we have*

$$closure(M) = models(\Sigma)$$

Proof. By Theorem 2 there is some Horn theory Σ' whose set of models is $closure(M)$. Plainly Σ' is a Horn upper-bound. Now we claim that Σ' must be equivalent to Σ. Since again by Theorem 2, $models(\Sigma)$ is a closed set, and contains M, it also contains $closure(M)$. Therefore Σ has either the same models as Σ', or more models. But by definition Σ is the upper-bound with the fewest models, and is thus equivalent to Σ'. $\Box$

Thus the Horn approximation "weakens" the input data by adding in all the models generated by taking intersections. In our example above, we have that Σ_0 with models M_0' is the Horn approximation of M_0. Note that it is also the Horn approximation of, for example, $M_0 \cup \{[0100]\}$. When M is equal to its own closure, then it follows that the Horn approximation Σ is an *exact* fit to the data.

Next, we define the notion of a *characteristic* model. The characteristic models of a closed set M can be thought of as a minimal "basis" for M, that is, a smallest set that can generate all of M by taking intersections. In general, the characteristic models of any finite M are defined as those elements of M that do not appear in the closure of the rest of M:

Definition 4 Characteristic Model

Where M is a finite set of models, the set of characteristic models is given by

$$char(M) \stackrel{\text{def}}{=} \{m \in M \mid m \notin closure(M - \{m\})\}$$

For example, the characteristic models of M_0' are [1110], [1000], and [0101]. The other two models in M_0' can be obtained from these characteristic models via intersection.

According to the definition above the characteristic elements of any set of models are unique and well-defined. Furthermore, the characteristic models of a set can generate the complete closure of the set. Now, because the set of models of a Horn theory is closed, it follows that we can identify a Horn theory with just the characteristic elements among its models. (In fact, henceforth we will simply say "the characteristic models of a Horn theory" to mean the characteristic subset of its models.) In general, this set may be much smaller than the set of all of its models. Finally, we arrive at an alternative characterization of the notion of a Horn approximation of a set M: as the closure of the set of characteristic models of M. The following theorem summarizes this discussion. Each property follows fairly directly from the above definitions.

Proposition 5 *Let M be any finite set of models. Then,*

(i) *$char(M)$ is unique,*
(ii) *$closure(char(M)) = closure(M)$,*
(iii) *If Σ is a Horn theory then $closure(char(models(\Sigma))) = models(\Sigma)$,*
(iv) *If Σ is the Horn approximation of M,*
then $closure(char(M)) = models(\Sigma)$.

Characteristic models are called "extreme" models in [5]. The proofs in this paper all depend upon the assumption that we are dealing with finite sets of models: certain infinite sets of models (over an infinite number of variables) may have no characteristic elements.

As an aside, one should note that the notion of a characteristic model is not the same as the standard definition of a maximal model. By definition, any $m \in M$ is a maximal model of M iff there is no $m' \in M$ such that the variables assigned to "true" by m' are a superset of those assigned to "true" by m. It is easy to see that all maximal models of a set (or theory) are characteristic, but the reverse does not hold. For example, the model [1000] in M_0 is an example of a non-maximal characteristic model.

3 Size of Representations

In this section we will examine the most concise way of representing the information inherent in a Horn theory. We have three candidates: a set of Horn clauses; the complete set of models of the theory; and the set of characteristic models of the theory. [3]

We can quickly eliminate the complete set of models from contention. Obviously, it is as least as large as the set of characteristic models, and often much larger. Furthermore, every Horn theory with K models over n variables can be represented using at most Kn^2 Horn clauses [5]. Thus up to a small polynomial factor, the complete set of models is also always at least as large as the clausal representation.

Neither of the other two representations strictly dominates the other. We first show that in some cases the representation using characteristic models can be exponentially smaller than the *best* representation that uses Horn clauses.

Theorem 6 *There exist Horn theories with $O(n^2)$ characteristic models where the size of the smallest clausal representation is $O(2^n)$.*

Proof. Consider the theory $\Sigma = \{\neg x_1 \vee \neg x_2 \vee \ldots \vee \neg x_n | \; x_i \in \{p_i, q_i\}\}$. The size of Σ is $O(2^n)$. Moreover, we show that there is no shorter clausal form for Σ, but the size of its set of characteristic models is polynomial in n.

Observe that no two clauses in Σ resolve. First we will prove that Σ is irredundant (no subset of Σ implies all of Σ). Then we will prove that Σ is of minimum size. (Note that being of minimum size is a stronger condition than being irredundant.)

Proof that Σ is irredundant: suppose there is a clause α in Σ such that $\Sigma - \{\alpha\} \models \alpha$. Since no two clauses in $\Sigma - \{\alpha\}$ resolve, by completeness of resolution there must be an α' in $\Sigma - \{\alpha\}$ such that α' subsumes α. But this is impossible, since all clauses in Σ are of the same length.

Next, we prove that there is no smaller set of clauses Σ' which is equivalent to Σ: Suppose there *were* an Σ' such that $\Sigma \equiv \Sigma'$ and $|\Sigma'| < |\Sigma|$. Then for all

[3] Another representation one could consider is DNF (disjunctive normal form). However, R. Khardon and D. Roth have recently shown that the characteristic model representation is more concise than DNF. Specifically, the number of characteristic models of a Horn theory is bounded by the size of the minimal DNF representation times the number of variables; on the other hand, the minimal DNF may be exponentially larger than the number of characteristic models [12].

α in Σ', it is the case that $\Sigma \models \alpha$. Because no clauses in Σ resolve, this means that there exists an α' in Σ such that α' subsumes α.

That is, every clause in Σ' is subsumed by some clause in Σ. Suppose each clause in Σ subsumed a different clause in Σ'; then $|\Sigma'| \geq |\Sigma|$, a contradiction. Therefore there is a proper subset Σ'' of Σ such that each clause in Σ' is subsumed by some clause in Σ''.

Then $\Sigma'' \models \Sigma'$, and therefore $\Sigma'' \models \Sigma$. But this is impossible, because we saw that Σ is irredundant. Therefore there is no smaller set of clauses equivalent to Σ which is shorter than Σ.

Thus we have shown that the smallest clausal representation of the set of models described by Σ is just Σ itself, and thus is exponential in terms of the number of variables n. Now we show that the set of characteristic models that describes this theory is polynomial in n.

Write a model as a truth assignment to the variables $p_1q_1p_2q_2 \ldots p_nq_n$. From the clauses in Σ, it is clear that in each model there must be some pair p_i and q_i where both letters are be assigned false (otherwise, there is always some clause eliminating the model). Without loss of generality, let us consider the set of models where p_1 and q_1 are both assigned false. Each of the clauses in Σ is now satisfied, so we can set the other letters to any arbitrary truth assignment. The characteristic models of this set are

$$\begin{array}{lcl}
[00111111\ldots11] & & [00111111\ldots11] \\
[00011111\ldots11] & \ldots\ldots & [00111111\ldots01] \\
[00101111\ldots11] & & [00111111\ldots10]
\end{array}$$

The three models in the first column represent all the settings of the second pair of letters. (Note that 00 can be obtained by intersecting the 2nd and the 3rd model.) Each triple handles the possible settings of one of the pairs. From these $3(n-1)$ models, we can generate via intersections all possible truth assignments to the letters in all pairs other than the first pair. For each pair, we have a similar set of models with that pair set negatively. And, again each set can be generated using $3(n-1)$ models. So, the total number of characteristic models is at most $O(n^2)$. $\Box$

The following theorem, however, shows that in other cases, the set of characteristic models can be exponentially *larger* than the best equivalent set of Horn clauses.

Theorem 7 *There exist Horn theories of size $O(n)$ with $O(2^{(n/2)})$ characteristic models.*

Proof. Consider the theory Σ given by the clauses $(\neg a \vee \neg b)$, $(\neg c \vee \neg d)$, $(\neg e \vee \neg f)$, etc. The set M of characteristic models of this theory contains all the models where each of the variables in each consecutive pair, such as (a, b), (c, d), (e, f), etc., are assigned opposite truth values (*i.e.*, either [01] or [10]). So, we get the models [010101...], [100101...], [011001...], ..., [101010...]. There are $2^{(n/2)}$ of such such models, where n is the number of variables. It is easy to see that these are all maximal models of the theory, and as we observed earlier, all such models are characteristic. (One can go on to argue that there are no other characteristic models in this case.) $\square$

Thus we see that sometimes the characteristic model set representation offers tremendous space-savings over the clausal representation, and vice-versa. This suggests a strategy if one wishes to compactly represent the information in a closed set of models: interleave the generation of both representations, and stop when the smaller one is completed.

The characteristic models in a closed set can be efficiently found by selecting each model which is not equal to the intersection of any two models in the set. The clausal theory can be found using the algorithms described in [5] and [11]. We will return to the problem of generating efficient representations in Section 6 below.

4 Deduction using Characteristic Models

One of the most appealing features of Horn theories is that they allow for fast inference. In the propositional case, queries can be answered in polynomial time [6]. However, there is no *a priori* reason why a representation based on characteristic models would also enable fast inference. Nevertheless, in this section, we show that there is indeed a polynomial-time algorithm for deduction using characteristic models.

We will take a query to be a formula in conjunctive normal form — that is, a conjunction of clauses. It is easy to determine if a query follows from a complete set of models: simply verify that the query evaluates to "true" on every model. But if the representation is just the set of characteristic models, such a simple approach does not work. For example, let the query α be the formula $a \vee b$, and let the characteristic set of models be $M_0 = \{[1110], [0101], [1000]\}$, as defined earlier. It is easy to see that α evaluates to true in each member of M_0. However, α does not logically follow from the Horn theory with characteristic model set M_0; in other words, α does not hold in every model in the closure of M_0. For example, the query is false in $[0101] \cap [1000] = [0000]$.

There is, however, a more sophisticated way of evaluating queries on the set of characteristic models, that does yield an efficient sound and complete algorithm. Our approach is based on the idea of a "Horn-strengthening", which we introduced in [21,22].

Definition 8 Horn-strengthening
A Horn clause C_{H} is a Horn-strengthening of a clause C iff C_{H} is a Horn clause, C_{H} subsumes C, and there is no other Horn clause that subsumes C and is subsumed by C_{H}.

Another way of saying this is that a Horn-strengthening of a clause is generated by striking out positive literals from the clause just until a Horn clause is obtained. For example, consider the clause $C = p \vee q \vee \neg r$. The clauses $p \vee \neg r$ and $q \vee \neg r$ are Horn-strengthenings of C. Any Horn clause has just one Horn-strengthening, namely the clause itself.

A key property of Horn theories is described by the following lemma.

Lemma 9 *Let Σ_{H} be a Horn theory and C a clause that is not a tautology. If $\Sigma_{\mathrm{H}} \models C$ then there is a clause C_{H} that is a Horn-strengthening of C such that $\Sigma_{\mathrm{H}} \models C_{\mathrm{H}}$.* $\square$

Proof. By the subsumption theorem [15], there is a clause C' that follows from Σ_{H} by resolution such that C' subsumes C. Because the resolvent of Horn clauses is Horn, C' is Horn. Either C' itself is a Horn-strengthening of C (so $C' = C_{\mathrm{H}}$),or C' subsumes some Horn-strengthening C_{H}. In either case, $\Sigma_H \models C' \models C_{\mathrm{H}}$. $\square$

Suppose the query is a single clause. Then the following theorem shows how to determine if the query follows from a knowledge base represented by a set of characteristic models.

Theorem 10 *Let Σ be a Horn theory and M its set of characteristic models. Further let C be any clause. Then $\Sigma \models C$ iff there exists some Horn-strengthening C_{H} of C such that C_{H} evaluates to "true" in every model in M.*

Proof. Suppose $\Sigma \models C$. By Lemma 9, $\Sigma \models C_{\mathrm{H}}$ for some Horn-strengthening C_{H} of C. So C_{H} evaluates to "true" in every model of Σ, and thus in every member of M. On the other hand, suppose that there exists some Horn-strengthening C_{H} of C such that C_{H} evaluates to "true" in every model in M. By Theorem 2, because the elements of M are models of a Horn theory C_{H}, the elements of the closure of M are all models of C_{H}. But the closure of M is the models of Σ; thus $\Sigma \models C_{\mathrm{H}}$. Since $C_{\mathrm{H}} \models C$, we have that $\Sigma \models C$. $\square$

In the previous example, one can determine that $a \vee b$ does not follow from the theory with characteristic models M_0 because neither the Horn-strengthening a nor the Horn-strengthening b hold in all of $\{[1110], [0101], [1000]\}$.

A clause containing k literals has at most k Horn-strengthenings, so one can determine if it follows from a set of characteristic models in k times the cost of evaluating the clause on each characteristic model. In the more general case the query is a conjunction of clauses. Such a query can be replaced by a sequence of queries, one for each conjunct. We therefore obtain the following theorem:

Theorem 11 *Let a Horn theory Σ be represented by its set of characteristic models M, and let α be a formula in conjunctive normal form. It is possible to determine if $\Sigma \models \alpha$ in time $O(n|M||\alpha|^2)$, where n is number of variables.*

5 Abduction using Characteristic Models

Another central reasoning task for intelligent systems is abduction, or inference to the best explanation [19]. In an abduction problem, one tries to *explain* an observation by selecting a set of assumptions that, together with other background knowledge, logically entails the observation. This kind of reasoning is central to many systems that perform diagnosis or interpretation, such as the ATMS.

The notion of an explanation can be formally defined as follows [20]:

Definition 12 [Explanation] *Given a set of clauses Σ, called the background theory, a subset A of the propositional letters, called the assumption set, and a query letter q, an explanation E for q is a minimal subset of unit clauses with letters from among A such that*

1. $\Sigma \cup E \models q$, *and*
2. $\Sigma \cup E$ *is consistent.*

Note that an explanation E is a set of unit clauses, or equivalently, a single conjunction of literals.

For example, let the background theory be $\Sigma = \{a, \ \neg a \vee \neg b \vee \neg c \vee d\}$ and let the assumption set $A = \{a, b, c\}$. The conjunction $b \wedge c$ is an explanation for d.

It is obvious that in general abduction is harder than deduction, because the definition involves both a test for entailment and a test for consistency. However, abduction can remain hard even when the background theory is restricted to languages in which both tests can be performed in polynomial time. Selman and Levesque [23] show that computing such an explanation is NP-complete even when the background theory contains only Horn clauses, despite the fact that the tests take only linear time for such theories. The problem remains hard because all known algorithms have to search through an exponential number of combinations of assumptions to find an explanation that passes both tests.

There are very few restricted clausal forms for which abduction is tractable. One of these is *definite* Horn clauses, which are Horn clauses that contain exactly one positive literal — completely negative clauses are forbidden. However, the expressive power of definite Horn is much more limited than full Horn: In particular, one cannot assert that two assumptions are mutually *incompatible.*

It is therefore interesting to discover that abduction problems can be solved in polynomial time when the background theory is represented by a set of characteristic models. We give the algorithm for this computation in Fig. 2. Note that the algorithm takes advantage of the fact that when the background theory Σ is Horn, every explanation contains only positive literals (*i.e.,* each explanation is simply a subset of A).

The abduction algorithm works by searching for a characteristic model in which the query holds. Then it sets E equal to the strongest set of assumptions that is compatible with the model, and tests if this E rules out all models of the background theory in which the query does not hold. This step is performed by the test

$$closure(M) \models (\bigwedge E) \supset q$$

and can be performed in polynomial time, using the deduction algorithm described in the previous section. (Note that the formula to be deduced is a single Horn clause.) If the test succeeds, then the assumption set is minimized, by deleting unnecessary assumptions. Otherwise, if no such characteristic model is in the given set, then no explanation for the query exists. Note that the minimization step simply eliminates redundant assumptions, and does not try to find an assumption set of the smallest possible cardinality, so no combinatorial search is necessary.

It is easy to see that if the algorithm does find an explanation it is sound, because the test above verifies that the query follows from the background theory together with the explanation, and the fact that the model m is in

```
function Explain(M, A, q)
    for each m in M do
        if m ⊨ q then
            E ← all letters in A that
                are assigned "true" by m
            if closure(M) ⊨ (⋀E) ⊃ q then
                Minimize E by deleting as many
                    elements as possible while
                    maintaining the condition
                    that closure(M) ⊨ (⋀E) ⊃ q.
                return E
            endif
        endif
    endfor
    return "false"
end.
```

Fig. 2. Polynomial time algorithm for abduction. M is a set of characteristic models, representing a Horn theory; A is the assumption set; and q is the letter to be explained. The procedure returns a subset of A, or "false", if no explanation exists.

M (and thus also in the closure of M) ensures that the background theory and the explanation are mutually consistent. Furthermore, if the algorithm searched through *all* models in the closure of M, rather than just M itself, it would be readily apparent that the algorithm is complete. (The consistency condition requires that the the explanation and the query both hold in at least one model of the background theory.) However, we will argue that it is in fact only necessary to consider the *maximal* models of the background theory; and since, as we observed earlier, the maximal models are a subset of the characteristic models, the algorithm as given is complete.

So suppose m is in $closure(M)$, and E is a subset of A such that q and all of E hold in m. Let m' be any maximal model of M (and thus, also a maximal model of $closure(M)$) such that $m \sqsubseteq m'$ — at least one such m' must exist. All the variables set to "true" in m are also set to "true" in m'; and furthermore, q and all of E consist of only *positive* literals. Therefore, q and E both hold in m' as well.

Thus the algorithm is sound and complete. In order to bound its running time, we note that the outer loop executes at most $|M|$ times, the inner (minimizing) loop at most $|A|$ times, and each entailment test requires at most $O(n|M||A|^2)$ steps. Thus the overall running time is bounded by $O(n|M|^2|A|^3)$. In summary:

Theorem 13 *Let M be the set of characteristic models of a background Horn theory, let A be an assumption set, and q be a query. Then one can find an abductive explanation of q in time $O(n|M|^2|A|^3)$.*

The fact that abduction is hard for clausal Horn theories, but easy when the same background theory is represented by a set of characteristic models, means that it may be difficult to generate the characteristic models of a given Horn theory: there may be exponentially many characteristic models, or even if there are few, they may be hard to find. None the less, it may be worthwhile to invest the effort to "compile" a useful Horn theory into its set of characteristic models, just in case the latter representation does indeed turn out to be of reasonable size. This is an example of "knowledge compilation" [21,22], where one is willing to invest a large amount of off-line effort in order to obtain fast run-time inference. Alternatively, one can circumvent the use of a formula-based representation all together by constructing the characteristic models by hand, or by generating them from empirical data, aas described in the next section.

6 Generating Characteristic Model Approximations

So far, we have compared the size of model-based representations to that of clausal representations, and compared their computational properties. We now turn our attention to the question of how to obtain characteristic model or clausal representations when given as input a a set of models (or cases, as discussed in the introduction).

When we find that the set of models of a Horn approximation of a set of input models is *identical* to the input set, we say that we have identified a Horn theory. Using the closure property of Horn theories it is easy to determine whether the input set corresponds to a Horn theory: For each pair of models in the input set, determine whether the intersection of those models is also in the input set. The complexity of this procedure is $O((n|M|)^2)$. More interestingly, Dechter and Pearl [5] show that if a set of input models M does correspond to a Horn theory, then this theory can be represented using at most $n^2|M|$ clauses, and the clauses can be generated in polynomial time.

When the set of input models does not correspond to a Horn theory, we want to find the Horn approximation of that set of models. The characteristic models from a (not necessarily closed) set M can be selected by testing for each $m \in M$ whether $m \in \mathit{closure}(M)$, using Definition 4; by Proposition 5, this gives the characteristic model representation of the Horn approximation of M. A naïve implementation of this algorithm, however, would require exponential time, because the closure of $M - \{m\}$ may be exponentially larger than M. Fortunately, it is possibly to efficiently check whether a model falls in the closure of a set of models, without actually generating that closure.

The algorithm to perform this test appears in Fig. 3. It is based on the ob-

function inClosure(m, M)
    $M' := \{m' | m' \in M \text{ and } m \sqsubseteq m'\}$
    if $M' = \emptyset$ then return "false" endif
    if $m = (\bigcap M')$ then return "true"
    else return "false"
    endif
end.

Fig. 3. Algorithm to test membership of a model m in the closure of a set of models M. It returns "true" if $m \in \mathit{closure}(M)$, and "false" otherwise.

servation that in trying to generate a model m by taking intersections, one need only consider models that assign true to all variables assigned true by m. Taking an intersection with any m' for which this did not hold would result in a model that assigned false to at least one variable that m assigned true. Furthermore, in trying to generate m one may as well intersect all models that assign true to the variables assigned true by m. Even if only a subset of these models generates m, intersecting with the other models still results in m.

The *inClosure* algorithm runs in $O(n|M|)$. Therefore computing the characteristic model representation of an arbitrary set of models can be done in $O(n|M|^2)$ time.

7 Computing General Horn Approximations

We have seen that the characteristic model representation of the Horn approximation of a set of models can be computed and reasoned with efficiently. None the less, for some applications there may be other reasons to prefer a rule-like, clausal representation. Therefore let us consider the problem of generating a clausal representation of the Horn approximation of a set of models.

It follows from Theorem 6 that the clausal representation of the Horn approximation of M can be exponentially large. Dechter and Pearl [5] therefore investigated k-Horn approximations. The maximum size of a k-Horn approximation is given by the maximum number of distinct Horn clauses with at most k literals, which is polynomial in n. Dechter and Pearl give a polynomial time algorithm for generating such k-Horn approximation. Of course, intuitively speaking, k-Horn approximations may not be as good as some less restricted Horn approximation. In fact, the following proposition shows that a k-Horn approximations can be very bad compared to even an only slightly more general Horn approximation, such as a $(k+1)$-Horn approximation.

Theorem 14 *For any k, there exists a set of models of size $O(n^{k+1})$, where the best k-Horn approximation has $O(2^n)$ models, whereas the $(k+1)$-Horn* *approximation has $O(n^{k+1})$ models.*

Proof. We consider a $(k+1)$-Horn theory that has no good k-Horn approximations. Let $S = \{x_1, \ldots, x_n\}$ be the set of propositional variables. Let Σ be the $(k+1)$-Horn theory

$$\bigwedge_{\{y_1,\ldots,y_{k+1}\} \subseteq S} (\neg y_1 \vee \neg y_2 \vee \ldots \neg y_{k+1}),$$

The models of Σ are exactly those assignments with at most k variables set to true. There are roughly $O(n^{k+1})$ such models total, so certainly the set of characteristic models is bounded in size by $O(n^{k+1})$.

Now, consider any k-Horn clause, say $(\neg x_1 \vee \ldots \vee \neg x_k)$. If we set the variables it contains to true, then we have falsified the clause, but have set at most k variables to true. Thus we can still extend our assignment to a model of Σ. Therefore, the best k-Horn theory approximating Σ is the empty theory which has 2^n assignments as models. $\square$

Given Theorem 14, it is clear that it can still be a good idea to generate unrestricted Horn approximations. Therefore we will present an algorithm for generating a general clausal presentation of the Horn approximation of a given input model set.[4] It is unknown whether there exists a exact, polynomial, deterministic algorithm for this task. However, we can come close:

Theorem 15 *Let Σ be the smallest Horn theory such that $models(\Sigma) = closure(M)$. There is a randomized algorithm that takes as input both M and small real numbers ϵ and δ (where $0 < \epsilon,\ \delta \leq 1$), and outputs a Horn theory $\hat{\Sigma}$ such that with probability $1 - \delta$,*

$$closure(M) = models(\Sigma) \subseteq models(\hat{\Sigma})$$

[4] Dechter and Pearl [5] observe that when the closure of the input set M is reasonably small ($|closure(M)|/|M|$ is bounded by some constant), general Horn approximations can also be computed by simply generating the complete closure and then directly applying the algorithm for the exact case. The randomized algorithm described below, however, can be used even when the closure is large. For example, suppose $|M| = n$, and $|closure(M)| = 2^{(n/2)}$. In this case it is impractical to generate the closure. However, the fraction of all models on which M and its Horn approximation disagree is only $(2^{(n/2)} - n)/2^n \approx 1/2^{(n/2)}$, that is, vanishingly small. Thus, even in this case it may be quite desirable to generate a representation of the Horn approximation.

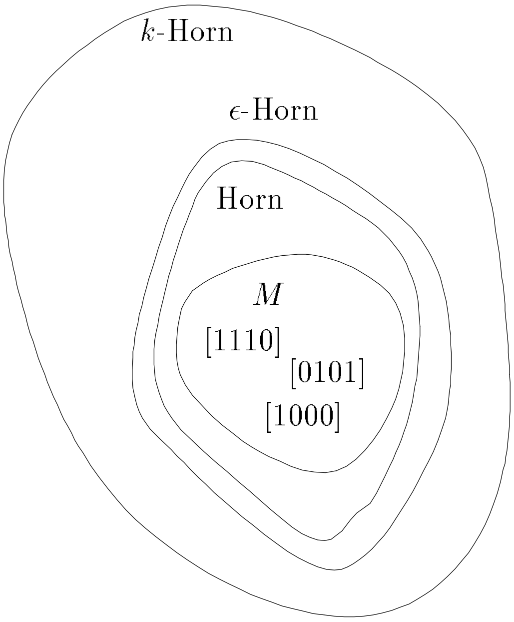

Fig. 4. Relative size of the sets of models contained in M, the general Horn approximation of M, the ϵ-good Horn approximation, and the $k-$Horn approximation.

and

$$\frac{|models(\hat{\Sigma}) - models(\Sigma)|}{2^n} \leq \epsilon.$$

This algorithm runs in time $poly(n, |M|, |\Sigma|, \frac{1}{\epsilon}, \frac{1}{\delta})$.

Thus, $\hat{\Sigma}$ will be an overgeneralization of the desired Horn approximation Σ, but it is a *controllable* overgeneralization: for any small ϵ, we can ensure that the fraction of all assignments that are models of $\hat{\Sigma}$ but not models of Σ is smaller than ϵ. See Fig. 3.

Our algorithm is based on an algorithm for learning Horn theories from examples (*i.e.*, models) by Angluin *et al.* [1]. Angluin *et al.*'s algorithm is quite involved, and we refer the reader to their paper for the details. The algorithm employs a *membership oracle* and an *equivalence oracle* to construct a formula that is logically equivalent to the unknown Horn formula that is to be learned. Given a truth assignment the *membership oracle* determines whether or not the assignment is a model of the formula to be learned. The *equivalence oracle* takes as input a theory and determines whether it is logically equivalent to the formula to be learned. If that is not the case, the oracle returns a counterexample, *i.e.*, a truth assignment on which the given theory differs from the unknown one. Angluin's *et al.*'s algorithm runs in time polynomial in the number of clauses and variables used in the unknown formula Σ.

The basic idea behind our approach is to replace the oracles by polynomial time deterministic procedures that use the model set M. First, to simulate a membership query m for Σ, we must test whether $m \in closure(M)$. This can be done in $O(n|M|)$ time using the *inClosure* algorithm described earlier (Fig. 3).

To simulate an equivalence query for Σ, we must somehow be able to efficiently determine if $models(\Sigma') = closure(M)$ using only the model set M, where Σ' is a Horn theory conjectured by the simulation of the Angluin *et al.* algorithm. For some Σ' this can be done easily: namely, if $closure(M) \not\subseteq models(\Sigma')$. This means that there must be a model $m \in M$ that is not a model of Σ', and we can detect this by a simple scan of M. Thus, our first step in simulating an equivalence query Σ' will be to make sure that all models in M are models of Σ'. If not, we have a counterexample for Σ', and the equivalence query is complete. If so, then we know that $closure(M) \subseteq models(\Sigma')$ (that is, Σ' is an overgeneralization).

In the next step we need to determine whether Σ' has some model that is not in $closure(M)$. If we had an efficient way of listing the characteristic models for a given Horn theory, we could check that each characteristic model of Σ' is in M. As soon as we encounter characteristic model that is not in M, we know that Σ' contains strictly more models than $closure(M)$. Unfortunately, as of yet, no efficient algorithm for generating characteristic models has been found. Kavvadias *et al.* [10] have recently shown that the problem is at least as hard as the so-called hypergraph enumeration problem, which is a well-known open problem [7].

We will therefore use a random sampling strategy to search for a possible counterexample for the overgeneralization Σ'. Without loss of generality, we assume that

$$\frac{|models(\Sigma') - models(\Sigma)|}{2^n} > \epsilon.$$

Note that if this condition is not satisfied, then Σ' already meets the criteria for the final Horn theory $\hat{\Sigma}$, and so we may stop and output $\hat{\Sigma} = \Sigma'$.

Under this condition, if we choose an assignment randomly then we have probability at least ϵ of drawing an m that is a model of Σ' but not of Σ and thus is a counterexample to Σ. It is easy to test whether we have drawn such an m: we first use the conjectured Horn theory Σ' (whose explicit representation is given to us by the learning algorithm), and check that m is a model of Σ'. We then use the test described above to check that m is not in $closure(M)$. If m meets both conditions, we have a counterexample. If one of the checks fails, we repeat the process with another randomly chosen assignment. We repeat this process a total of l times (per equivalence query); if all l tries fail, we simply output Σ'.

The probability that we fail to find a counterexample in l trials can be made smaller than δ' for $l = O(\frac{1}{\epsilon}\log\frac{1}{\delta'})$ by a simple and standard probabilistic analysis. If we wish to achieve a global failure probability of at most δ, then we can set $\delta' = \frac{\delta}{Q}$, where Q is the total number of equivalence queries that can be made (a polynomial in all the relevant quantities). Thus, with probability at least $1-\delta$, this algorithm will output $\hat{\Sigma}$ with the claimed properties.

8 Conclusions

In this paper, we have demonstrated that, contrary to prevalent wisdom, knowledge-based systems can efficiently use representations based on sets of models rather than logical formulas. Incomplete information does not necessarily make model-based representations unwieldy, because it possible to store only a subset of characteristic models that are equivalent to the entire model set. We showed that for Horn theories neither the formula nor the model-based representation dominates the other in terms of size, and that sometimes one other can offer an exponential savings over the other. Recently, Khardon and Roth [12] have introduced an interesting generalization of our model-based representation, and have shown a clear computational advantage of the use of their model-based representation by combining various learning and reasoning tasks in a single framework [13].

We also showed that the characteristic model representation of Horn theories has very good computational properties, in that *both* deduction and abduction can be performed in polynomial time. On the other hand, all known and foreseeable algorithms for abduction with Horn clauses are of worst-case exponential complexity.

We concluded by examining algorithms for generating efficient representations of the Horn approximation of a set of models. In particular, we presented a randomized algorithm for computing a general clausal representation of the Horn approximation of an arbitrary model set.

Acknowledgement

We thank Rina Dechter and Judea Pearl for many interesting discussions and comments.

References

[1] D. Angluin, M. Frazier, and L. Pitt, Learning conjunctions of Horn clauses, in: *Proceedings 31st Annual Symposium on Foundations of Computer Science*, St. Louis, MI (IEEE Computer Society Press, Los Almamitos, CA, 1990) 186–192.

[2] C. Beeri, M. Dowd, R. Fagin, and R. Statman, On the structure of Armstrong relations for functional dependencies, *J. ACM* **31**(1) (1984) 30–46.

[3] T. Bylander, D. Allemang, M.C. Tanner, and J.R. Josephson, Some results concerning the computational complexity of abduction, in: *Proceedings KR-89*, Toronto, Ontario (1989) 44–54.

[4] R. Dechter and A. Itai, The complexity of finding all solutions, UCI Technical Report 92-61, University of California at Irvine, Irvine, CA (1991).

[5] R. Dechter and J. Pearl, Structure identification in relational data, *Artif. Intell.* **58** (1992) 237–270.

[6] W.F. Dowling and J.H. Gallier, Linear time algorithms for testing the satisfiability of propositional Horn formula, *Journal of Logic Programming* **3** (1984) 267–284.

[7] T. Eiter and G. Gottlob, Identifying the minimal traversals of hypergraph and related problems, preprint (1991).

[8] K. Eshghi, A tractable class of abduction problems, in: *Proceedings IJCAI-93*, Chambéry, France (1993) 3–8.

[9] A. Horn, On sentences which are true of direct unions of algebras, *Journal of Symbolic Logic* **16** (1951) 14–21.

[10] D. Kavvadias, C. Papadimitriou, and M. Sideri, On Horn envelopes and hypergraph traversals, *Proceedings Fourth Annual International Symposium on Algorithms and Computation (ISAAC-93)*, Hong Kong (1993).

[11] H. Kautz, M. Kearns, and B. Selman, Reasoning with characteristic models, *Proceedings AAAI-93*, Washington, DC (1993) 34–39.

[12] R. Khardon and D. Roth, Reasoning with models, Tech. Report TR-1-94, Center for Research in Computing Technologies, Harvard University, Boston, MA (1994). A short version appears in: *Proceedings AAAI-94*, Seattle, WA (1994) 1148–1153.

[13] R. Khardon and D. Roth, Learning to reason, in: *Proceedings AAAI-94*, Seattle, WA (1994) 682–687.

[14] J.L. Kolodner, *Case-Based Reasoning* (Morgan Kaufmann, San Mateo, CA, 1993).

[15] R.C.T. Lee, *A completeness theorem and a computer program for finding theorems derivable from given axioms*, PhD thesis, University of California at Berkeley, Berkeley, CA (1967).

[16] H. Manilla and K.J. and Raiha, *Design of Relational Databases* (Addison-Wesley, Reading, MA, 1992).

[17] J. McCarthy and P.J. Hayes, Some philosophical problems from the standpoint of artificial intelligence, in: D. Michie, ed., *Machine Intelligence 4* (Ellis Horwood, Chichester, England 1969) 463–502.

[18] J.C.C. McKinsey, The decision problem for some classes of sentences without quantifiers, *Journal of Symbolic Logic* **8** (3) (1943) 61–77.

[19] C.S. Peirce, *Collected Papers of Charles Sanders Peirce* (Harvard University Press, Cambridge, MA 1958).

[20] R. Reiter and J. de Kleer, Foundations of assumption based truth maintenance systems: preliminary report, in: *Proceedings AAAI-87*, Seattle, WA (1987) 183–188.

[21] B. Selman and H. Kautz, Knowledge compilation using Horn approximations, in: *Proceedings AAAI-91*, Anaheim, CA (1991) 904–909.

[22] B. Selman and H. Kautz, Knowledge compilation and theory approximation *J. ACM*, in press.

[23] B. Selman and H.J. Levesque, Abductive and default reasoning: a computational core, in: *Proceedings AAAI-90*, Boston, MA (1990) 343–348.

[24] B. Selman and H.J. Levesque, Support set selection for abductive and default reasoning, *Artif. Intell.*, in press.