知識工学と人工知能 63 − 3
(1 9 8 9. 3. 1 3)

モデル論的考察による ATMS の一般化

井上 克巳

(財) 新世代コンピュータ技術開発機構
研究所 第 5 研究室

junet: inoue@icot.junet

概要

本論文では de Kleer の ATMS 及びその一般化に対する論理的基礎について議論する。まず、ATMS が本来持つべき要請を満足するように一般化を考える。次に、仮説推論のモデル論的考察に従って、拡張された ATMS に関する幾つかの性質を論じる。更に、この拡張された ATMS を基にして、(1) モデルを直接操作する「モデルに基づく ATMS」、(2) CWA を導入した「非単調 ATMS」、及び (3) 従来の研究で行われている幾つかの拡張例に関する論理的 (再) 構築を行う。

Generalizing the ATMS: A Model-based Approach

(Preliminary Report)

Katsumi Inoue

ICOT Research Center
1–4–28, Mita, Minato-ku, Tokyo 108, Japan

csnet: inoue%icot.jp@relay.cs.net
uucp: {kddlab,mit-eddie,ukc,enea,inria}!icot!inoue

Abstract

This paper describes logical specifications of de Kleer's assumption-based truth maintenance system (ATMS) and its generalizations based on principles which should be satisfied. A proposed generalized ATMS is investigated model-theoretically in terms of a hypothetical theory. A model-based ATMS, a nonmonotonic ATMS and other various generalizations of the ATMS are also explained by this general formalism.

1. Introduction

An *assumption-based truth maintenance system* (ATMS) [de Kleer 86] has been widely used when problems require reasoning in multiple contexts. However, this basic ATMS is restricted to accepting only Horn clause justifications and atomic assumptions. Until now, various generalizations have been proposed, such as [Reiter & de Kleer 87], [Dressler 87] and [de Kleer 88], but they have independently strengthened some features of the basic ATMS and are pursued algorithmically or syntactically rather than model-theoretically, so that it is not obvious how to make them more precise. The motivation of this research was the desire to formalize generalizations of the ATMS within a common simple model theory. We begin by introducing three postulates that should be satisfied by generalizations of the ATMS to be analyzed by this paper, as follows.

1. *Consistency maintenance in multiple contexts:* An ATMS should be defined as a *propositional hypothetical theory* composed of two sets of logical formulas: justifications (or *axioms*) and assumptions (or *hypotheses*). It is desirable for both sets to be extended to contain any propositional formulas. It is essential to maintain *consistency* of such a theory allowing for a concurrent representation of all contexts, and an ATMS would often be required to find maximal consistent solutions, or *extensions*.

2. *Abductive reasoning:* An ATMS should be able to generate multiple *sound* and *complete* explanations for a logical formula with respect to a hypothetical theory. Moreover, we often need *minimal* explanations, for efficiency, the principle of parsimony, or other reasons. A connection between hypothetical reasoning like the ATMS and abductive reasoning has been already discussed in [Reiter & de Kleer 87].

3. *Nonmonotonicity:* An ATMS which accepts any formula justification and assumptions should capture nonmonotonic reasoning because of introducing the notion of negation. In many cases, we want to supply *nonmonotonic justifications.*

2. General Formalism

We shall assume a set, $\mathcal{W}$, of *well-formed formulas* (wffs) constructed using a set, $\mathcal{A}$, of finitely many propositional symbols and logical connectives.[1] An *interpretation*, I, of $\mathcal{A}$ is defined as an element of $2^{\mathcal{A}}$ such that for each $\alpha \in I$, α is supposed to be assigned to true. The relations, $\models$ ($\subseteq 2^{\mathcal{A}} \times \mathcal{W}$; *satisfaction*) and $\models$ ($\subseteq 2^{\mathcal{W}} \times \mathcal{W}$; *entailment*), can be defined in the usual way. The set of all *models* of a set of wffs, W, is denoted as $M(W)$. For two wff sets, X and Y, $X \models Y$ is defined as $M(X) \subseteq M(Y)$. We say that $M \in M(W)$ is a *minimal* model of W iff $M' \in M(W)$ and $M' \subseteq M$ only if $M' = M$.

[1] While we use the propositional language to make the discussion clear, All concepts in this paper can be extended to have a subset of the first order predicate calculus without function symbol, where each formula is assumed to be universally quantified and domain closure axiom is included.

2.1 Supporting Hypotheses

Definition 1. A *hypothetical theory*, Δ, is a pair of W and D and is denoted as $\Delta = (W, D)$, where W is a satisfiable set of wffs, a set of *axioms*, and D is a set of wffs, a set of *hypotheses*. The set of *environments (with respect to Δ)* is defined as

$$Env(\Delta) = \{\, E \in 2^D \mid W \cup E \text{ is satisfiable} \,\}.$$

Suppose that $E \in Env(\Delta)$, then the *context of* (W, E) (denoted as $C(W, E)$) is the smallest set of wffs containing $W \cup E$ and closed under entailment. □

Definition 2. Let $w \in W$, and $\Delta = (W, D)$. An environment, $E \in Env(\Delta)$, is a *support for w (with respect to Δ)* iff

$$W \cup E \models w$$

holds. The set of all supports for w with respect to Δ is denoted as $S0(\Delta, w)$. □

Each support for w is a supplementary set of wffs with which w is entailed by W, keeping consistency with W. This perspective is simply the concurrent version of the framework in [Poole 87]. Given w, the computation of a set of supports for w corresponds to abductive reasoning. In this framework, we shall regard the notion of minimality in $S0(\Delta, w)$ as the *weakest* condition for explaining w as follows.

Definition 3. A support, $E \in S0(\Delta, w)$, is *(S1-)minimal for w (with respect to Δ)* iff

$$\neg \exists E' \in S0(\Delta, w).\ ((E \models E') \wedge (E' \not\models E)).$$

The set of all S1-minimal supports for w with respect to Δ is denoted as $S1(\Delta, w)$. □

Proposition 1. $S1(\Delta, w)$ has the following properties:
(1) for each $E \in S1(\Delta, w)$, $W \cup E$ is satisfiable,
(2) for each $E \in S1(\Delta, w)$, $w \in C(W, E)$,
(3) for $E \in Env(\Delta)$, $w \in C(W, E)$ iff $\exists E' \in S1(\Delta, w).\ (E \models E')$, and
(4) for every two environments, $E_1, E_2 \in S1(\Delta, w)$, $(E_1 \models E_2) \supset (E_2 \models E_1)$. □

In this model theory, the basic ATMS can be characterized as a hypothetical theory, $\Delta_{BATMS} = (H, A)$, where H is given as a set of Horn clauses and $A \subseteq \mathcal{A}$ is given as a set of atomic formulas, maintaining a concurrent representation of all contexts by labeling each atomic formula, a, with $S1(\Delta_{BATMS}, a)$ (called the *label* of a). Proposition 1 gives a natural generalization of the (1) *consistent*, (2) *sound*, (3) *complete* and (4) *minimal* properties of the ATMS labels, because we do not put any syntactical restriction on W and D. Proposition 1 (3) also characterizes a *membership problem:* given an environment, E, whether or not a given wff w holds in $C(\Delta, E)$. We characterize the notion of membership or minimality not by set inclusion of hypotheses but by what can be entailed. (Notice that for two formula sets, A, B, if $A \subseteq B$ then $B \models A$.) However, by considering more on models, we can give other criteria for minimality.

Definition 4. A support, $E \in S0(\Delta, w)$, is *S2-minimal for w (with respect to Δ)* iff

$$\neg\exists E' \in S0(\Delta, w).\ ((W \cup E \models E') \wedge (W \cup E' \not\models E)).$$

The set of all S2-minimal supports for w is denoted as $S2(\Delta, w)$. □

Example 1. Suppose that $\Delta_1 = (W, D_1)$ and $\Delta_2 = (W, D_2)$, where
$W = \{\, a \supset c,\ b \supset c,\ a \wedge b \supset g,\ \neg g \,\}$,
$D_1 = \{\, a,\ b,\ c \,\}$, and $D_2 = \{\, a,\ \neg g \supset a \,\}$.
From these, the following sets of supports can be obtained.
$S0(\Delta_1, c) = \{\, \{a\},\ \{b\},\ \{c\},\ \{a,\ c\},\ \{b,\ c\} \,\}$,
$S1(\Delta_1, c) = \{\, \{a\},\ \{b\},\ \{c\} \,\}$, $S2(\Delta_1, c) = \{\, \{c\} \,\}$,
$S0(\Delta_1, b \wedge c) = S2(\Delta_1, b \wedge c) = \{\, \{b\},\ \{b,\ c\} \,\}$, $S1(\Delta_1, b \wedge c) = \{\, \{b\} \,\}$,
$S1(\Delta_2, c) = \{\, \{\neg g \supset a\} \,\}$, and $S2(\Delta_2, c) = \{\, \{a\},\ \{\neg g \supset a\} \,\}$. □

Note that when hypotheses in D are independent of each other with respect to W, $S2(\Delta, w)$ is equivalent to $S1(\Delta, w)$. [2] To analyze the semantical properties of $Si(\Delta, w)$ $(i = 0, 1, 2)$, we define the set, $\mathcal{M}(W, S)$, of all models of W that satisfy all formulas in at least one element of a set, $S \subseteq 2^{W}$, of formula sets, as follows:

$$\mathcal{M}(W, S) = \bigcup_{E \in S} M(W \cup E)\,.$$

We define the set, $\mathcal{M}_0(\Delta, w)$, as: $\mathcal{M}_0(\Delta, w) = \mathcal{M}(W, S0(\Delta, w))$.

Theorem 2. $\mathcal{M}_0(\Delta, w) = \mathcal{M}(W, Si(\Delta, w))$ $(i = 1, 2)$. [3] □

Theorem 2 shows that $Si(\Delta, w)$ $(i = 1, 2)$ are equivalent to $S0(\Delta, w)$ in the sense that these sets have the same models of W that can satisfy w. Therefore, all the worlds where w holds can be characterized as $\mathcal{M}_0(\Delta, w)$ whatever minimality criterion is used.

2.2 Extensions

A set of hypotheses, D, is very closely related to a restricted case of a set of *normal defaults* in default logic [Reiter 80]. For each $d \in D$, d corresponds to the consequent of a normal default without a prerequisite, of the form :M $d\,/\,d$.

Definition 5. An *extension*, $\mathcal{E}$, *of* $\Delta = (W, D)$ is a context such that

$$\mathcal{E} = C(W,\ \{\, d \in D \mid \neg d \notin \mathcal{E} \})\,.$$

[2] Roughly speaking, an S2-minimal support corresponds to a *maximal assumption set* [Doyle 79].

[3] Proofs of all theorems appear in the full paper.

The *extension base for* $\mathcal{E}$ is defined as: $Eb(\mathcal{E}) = \{\, d \in D \mid \neg d \notin \mathcal{E} \}$. $\square$

All results of normal default theories in default logic, in particular the existence of extensions and *semi-monotonicity*, are guaranteed in our formalism.

Proposition 3. Let $\Delta = (W, D)$, and $\mathcal{E}(\Delta)$ be an extension of Δ.
(1) $Eb(\mathcal{E}(\Delta))$ is a maximal environment in $Env(\Delta)$.
(2) $M(\mathcal{E}(\Delta))$ is a minimal set of models in $\{\, M(W \cup E) \mid E \in Env(\Delta)\}$.
(3) For any wff, $w \in \mathcal{W}$, $w \in \mathcal{E}(\Delta)$ iff $Eb(\mathcal{E}(\Delta))$ is a maximal support in $\mathcal{S}0(\Delta, w)$. $\square$

2.3 Model-based ATMSs

Now, we propose an alternative way of maintaining all contexts by focusing on Herbrand interpretations of a set of hypotheses instead of keeping minimal supports for a wff.

Definition 6. Let $\Delta = (W, D)$. The ***hypothetical base***, $P(D) \subseteq \mathcal{A}$, *of* D is the set of all atomic formulas occurring in D. The set of all *h-interpretations (with respect to* Δ*)* is defined as $\mathcal{H}(\Delta) = \{\, M \cap P(D) \mid M \in M(W)\,\}$. For $I \in 2^{P(D)}$, $\bar{I}$ is defined as $\bar{I} = \{\neg p \mid p \in P(D) - I\}$. Let $w \in \mathcal{W}$. Then, the following two sets of h-interpretations satisfying w are defined as:

$$\Psi(\Delta, w) = \{\, M \cap P(D) \mid M \in \mathcal{M}_0(\Delta, w)\,\},$$

$$\Psi^*(\Delta, w) = \{\, I \in \mathcal{H}(\Delta) \mid W \cup I \cup \bar{I} \models w\,\}. \quad \square$$

We shall consider an ATMS handling $\Psi(\Delta, w)$ for w, called a *model-based ATMS*. While $\Psi^*(\Delta, w)$ can be computed from $\mathcal{H}(\Delta)$ restricted to the set of all h-interpretations that entail w, only parts of them are in $\Psi(\Delta, w)$ which is all that is needed for w to be realized by hypotheses, that is, $\Psi(\Delta, w) \subseteq \Psi^*(\Delta, w)$. However, there are sufficient conditions for the two sets to coincide. For a given set, D, of hypotheses, if all elements of $\{\, I \cup \bar{I} \mid I \in 2^{P(D)}\}$ can be considered to be realizable, the concept of model-based ATMSs is useful.[4] This case can be characterized as the following hypothetical theory.

$$\Delta_L = (W, L)\,, \quad \text{where } L \text{ is a set of } \textit{literals}, \text{ that is, } \ p \in L \equiv \neg p \in L\,.$$

Proposition 4. (1) $\Psi(\Delta_L, w) = \Psi^*(\Delta_L, w)$. (2) For each $I \in \Psi(\Delta_L, w)$, $I \cup \bar{I}$ is an extension base for an extension of Δ_L that contains w. $\square$

3. Nonmonotonicity

In this section, nonmonotonicity by an ATMS, in particular a model theory of *nonmonotonic justifications*, is investigated in our general formalism.[5]

[4] Several properties in which this condition is violated are given in the full paper.
[5] In the full paper, the treatment of *nonnormal defaults* [Reiter 80] is also discussed.

The *closed world assumption* (CWA) [Reiter 78] says that if a set, H, of Horn clauses does not entail an atomic formula, p, then $\neg p$ can be inferred. A generalization of the CWA for a set, W, of general wffs can be characterized in a similar way to [Reiter 80], as the *CWA hypothetical theory*:

$$\Delta_{CWA} = (W,\ \mathcal{N}_{\mathcal{A}}), \quad \text{where} \quad \mathcal{N}_{\mathcal{A}} = \{\, \neg p \mid p \in \mathcal{A}\,\}.$$

Then, a model of an extension of Δ_{CWA} is a minimal model of W.

We treat a *nonmonotonic* justification as a material implication in W of the form:

$$\alpha_1 \wedge \ldots \wedge \alpha_m \ \wedge \neg\beta_1 \wedge \ldots \wedge \neg\beta_n \ \supset \gamma, \quad \text{where} \quad \alpha_i, \beta_j, \gamma \in \mathcal{A},$$

and $\{\beta_j \mid j = 1, \ldots, n\}(\neq \phi)$ corresponds to the *outlist*. To enable simultaneous handling of nonmonotonic justifications and a set, A, of atomic hypotheses explicitly given to an ATMS, the above Δ_{CWA} can be extended as:

$$\Delta_{NM} = (W,\ A \cup \mathcal{N}_{\mathcal{A}}), \quad \text{where} \quad A \subseteq \mathcal{A}.$$

Because of the effect of introducing $\mathcal{N}_{\mathcal{A}}$, the membership by Proposition 1 (3) can be extended to the following problem: given an environment, $E \subseteq A \cup \mathcal{N}_{\mathcal{A}}$, whether or not a given wff w holds in an extension, $\mathcal{E}(\Delta_{NM}) = C(\Delta_{NM}, \tilde{E})$, where $\tilde{E}$ is a minimal environment with respect to Δ_{NM} (called a *complemented environment of* E) such that $\tilde{E} = E \cup \tilde{E}_N$, $\tilde{E}_N \subseteq \mathcal{N}_{\mathcal{A}}$. The next theorem characterizes this extended membership.

Theorem 5 (extended membership)**.** Suppose that E is an environment with respect to $\Delta_{NM} = (W,\ A \cup \mathcal{N}_{\mathcal{A}})$, where $E = E_P \cup E_N$, $E_P \subseteq A$, and $E_N \subseteq \mathcal{N}_{\mathcal{A}}$. Let $w \in \mathcal{W}$ and $\Delta_A = (W, A)$. Then, the following four statements are equivalent.
(1) There is a complemented environment, $\tilde{E}$, of E, such that $w \in C(\Delta_{NM}, \tilde{E})$.
(2) There is an S1-minimal support, $E' \in S1(\Delta_{NM}, w)$, where $E' = E'_P \cup E'_N$, $E'_P \subseteq A$, $E'_N \subseteq \mathcal{N}_{\mathcal{A}}$, such that $E_P \models E'_P$ and $W \cup E \cup E'_N$ is satisfiable [Dressler 87].
(3) There is a minimal model, M, of $W \cup E$ such that $M \models w$.
(4) There is a minimal h-interpretation, $I \in \Psi^*(\Delta_A, w)$, such that $W \cup E \models I$. □

Unfortunately, in the above discussion about nonmonotonicity by ATMSs, the notion of *groundedness* [Doyle 83] (or *well-foundedness*) cannot be incorporated, as our simple semantics with propositional logic cannot express it. To obtain groundedness, we must add some external mechanism to an ATMS. Therefore, an additional operational semantics is necessary to obtain expected models which reflect the unidirectional property or some intended meaning of justifications [Inoue 88].[6]

[6] The detailed discussion is given in the full paper.

4. Reconstruction of Various ATMSs

In Section 2.1, we characterized the basic ATMS as Δ_{BATMS}. This section shows how various generalizations of the ATMS can be reconstructed in our general formalism.

There are two recent reports on handling non-Horn clause justifications. The *negated assumption ATMS* (NATMS) [de Kleer 88] can be characterized as $\Delta_{NATMS} = (W, A)$, where $A \subseteq \mathcal{A}$ does not essentially treat negation hypotheses, and S1-minimal supports are computed for a literal. The label updating algorithm of the NATMS does not ensure the membership (Proposition 1 (3)), but does the extension completeness (Proposition 3 (3)). On the other hand, the *extended basic ATMS* (EATMS) [Dressler 87] can be characterized as $\Delta_{EATMS} = (W, A \cup \mathcal{N}_{\mathcal{A}})$, the same as Δ_{NM} in Section 3, and S1-minimal supports are computed for an atom. Dressler's 'Out-assumptions' and 'μ-extensions' correspond to our CWA and complemented environments in Theorem 5. Neither NATMS nor EATMS ensures well-foundedness as discussed in Section 3.

McDermott's nonmonotonic ATMS [McDermott 83] can be considered as a version of Δ_{NM} in Section 3, as it represents assumptions explicitly. McDermott proposed a constraint satisfaction method to compute minimal ('strictly OUTer') h-interpretations in Theorem 5 (4), but the method fails to capture groundedness.

The *clause management system* (CMS) [Reiter & de Kleer 87] can be characterized as $\Delta_{CMS} = (\Sigma, \mathcal{L}_{\mathcal{A}})$, where Σ is a set of clauses in clausal normal form and $\mathcal{L}_{\mathcal{A}}$ is the set of all literals, i.e., $\mathcal{L}_{\mathcal{A}} = \mathcal{A} \cup \mathcal{N}_{\mathcal{A}}$, and S1-minimal supports are computed for a clause. Δ_{CMS} is a special case of Δ_L in Section 2.3, and also satisfies Theorem 5.

The *massively parallel ATMS* (PATMS) [Dixon & de Kleer 88] is a realization of a model-based ATMS, and is characterized as $\Delta_{PATMS} = (W, L)$, the same as Δ_L in Section 2.3, and $\Psi(\Delta_{PATMS}, a)$ is computed for an atom, a, by Boolean constraint satisfaction methods. Hence, the PATMS satisfies Proposition 4.

Ginsberg's *first-order ATMS* [Ginsberg 88] handles closed formulas. He defines minimality in the way that an environment, E, is less minimal than another environment, E', iff for each wff, $w' \in E'$, there is a wff, $w \in E$, such that w is an instance of w'. This criterion is just the first-order version of our S1 minimality.

5. Conclusion

This paper presented a logical framework for a generalization of the ATMS in terms of a hypothetical theory. This is an important subcase of Reiter's default logic. The notion of minimality and completeness was extended and its model theory was discussed. Two important examples, a model-based ATMS and a nonmonotonic ATMS, were shown in the general framework. By using model-based ATMSs, the possibility of parallel label updating can be extended. Nonmonotonicity can be introduced in order to prune some environments that are not considered in commonsense. The CWA captures it partially. Various other generalizations of the ATMS were also re-examined by our formalism.

Acknowledgments

I would like to thank Johan de Kleer, Michael Reinfrank, Ulrich Junker, Ken Satoh, Jun Arima, and Koichi Furukawa for many helpful discussions.

References

[de Kleer 86] de Kleer, J., "An Assumption-based TMS", *Artificial Intelligence* **28** (1986), pp.127–162.

[de Kleer 88] de Kleer, J., "A General Labeling Algorithm for Assumption-based Truth Maintenance", *Proc. AAAI-88* (1988), pp.188–192.

[Dixon & de Kleer 88] Dixon, M. and de Kleer, J., "Massively Parallel Assumption-based Truth Maintenance", *Proc. AAAI-88* (1988), pp.199–204.

[Doyle 79] Doyle, J., "A Truth Maintenance System", *Artificial Intelligence* **12** (1986), pp.231–272.

[Doyle 83] Doyle, J., "The Ins and Outs of Reason Maintenance", *Proc. IJCAI-83* (1983), pp.349–351.

[Dressler 87] Dressler, O., *An Extended Basic ATMS*, Advanced Reasoning Methods, Report INF2 ARM-3-87, Siemens AG, 1987.

[Ginsberg 88] Ginsberg, M. L., "A Circumscriptive Theorem Prover: Preliminary Report", *Proc. AAAI-88* (1988), pp.470–474.

[Inoue 88] Inoue, K., *On the Semantics of Hypothetical Reasoning and Truth Maintenance*, ICOT Technical Report TR-357, ICOT, 1988.

[McDermott 83] McDermott, D., "Contexts and Data Dependencies: A Synthesis", *IEEE Trans. Pattern Anal. Machine Intelligence* **5** (3) (1983), pp.237–246.

[Poole 87] Poole, D. L., *A Logical Framework for Default Reasoning*, Department of Computer Science, Technical Report CS-87-59, University of Waterloo, 1987.

[Reiter 78] Reiter, R., "On Closed World Data Bases", in: *Logic and Data Bases* (Gallaire, H. and Minker, J., Eds.), Plenum Press, New York (1978), pp.55–76.

[Reiter 80] Reiter, R., "A Logic for Default Reasoning", *Artificial Intelligence* **13** (1980), pp.81–132.

[Reiter & de Kleer 87] Reiter, R. and de Kleer, J., "Foundations of Assumption-based Truth Maintenance Systems: Preliminary Report", *Proc. AAAI-87* (1987), pp.183–188.