TR-297

A Minimal Change of Belief-a Criterion of Belief Revision

by
K. Satoh

September, 1987

ICOT

Mita Kokusai Bldg. 21F
4-28 Mita 1-Chome
Minato-ku Tokyo 108 Japan

(03) 456-3191~5
Telex ICOT J32964

Institute for New Generation Computer Technology

極小変化による信念の修正について

佐藤 健

(財)新世代コンピュータ技術開発機構
ksatoh@icot.junet

知識が不完全な場合でも、日常生活では何らかの結論を下さなければならない場合がありうる。そうした場合に我々は自分の信念を使って知識の不足を補おうとする。しかし、もし新しく追加された知識が、その信念と矛盾していた場合には、信念を変化させて、その知識と矛盾しないようにしなければならない。本論文では、この信念の変化の基準として「極小変化」という考え方について述べる。「極小変化」とは、信念を変化させるときに、今まで信じていたことをできるだけ変化させないで、新しい知識と整合させることを定式化している。

A Minimal Change of Belief - a Criterion of Belief Revision

Ken Satoh

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

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

This paper presents a reasoning system of incomplete knowledge and a criterion of belief revision from the model theoretical viewpoint. If we want to draw a conclusion from incomplete knowledge, then we must use our belief to support any lack in the knowledge. However, when some added knowledge causes inconsistency with a current belief, we must change it to keep consistency. In this case, we must decide how the belief should be changed. This paper presents a criterion of belief revision called *minimal change*. *Minimal change* formalizes that we change our new belief to keep it as close to our previous belief as possible.

1. Introduction

In real life, we are usually faced with situations where we must draw a conclusion with incomplete knowledge. For one solution to those situations, we use our belief (or hypothesis) to complement unknown information. However, the belief is not always true, and when the belief or some information derived from it is found to be false, then we need to revise the belief in order to correct our knowledge.

The motivation of this research was to formalize the above situation, that is, to give simple and clear semantics of a reasoning mechanism with incomplete knowledge and belief, and to provide a criterion of belief revision. Several truth maintenance systems (TMSs) have been presented [Doyle79, Martins84, de Kleer86], but they mainly concentrate on efficient management in finding the origin of inconsistency (Doyle's TMS) or avoiding contradiction when assumptions are combined (Martin's MBR, de Kleer's ATMS). There has been some theoretical research on these systems [Doyle83, Martins86], but it is mainly based on syntactical analysis and the formalization of implementation. In [Reiter87], there is some model theoretic analysis of ATMS, but it is an analysis of aspects of hypothetical reasoning.

This paper presents a reasoning system of incomplete knowledge and a criterion of belief revision based on *minimal change. Minimal change* of belief means that belief is changed minimally when a contradiction occurs. When someone finds that his belief is contradictory, he does not want to throw away his entire belief, but to change it, as little as possible, to restore consistency. This situation is formalized in terms of the model theory.

2. Reasoning from Incomplete Knowledge with Belief

Belief should be satisfied under the following conditions.

(1) Beliefs can be combined to produce logical consequences.
(2) A belief does not contain contradictions.
(3) A belief must be consistent with known information (known information is called "*knowledge*").

Since this paper considers an ideal reasoner of incomplete knowledge and belief, condition (1) should hold. If the belief contains a contradiction, everything can be derived from the belief, so condition (2) should be satisfied. Condition (3) must be satisfied because inconsistency of belief with knowledge means incorrectness of belief. Taking these conditions into account, *knowledge* is defined here as propositional sentences and *belief* as a logical model of that *knowledge*. Propositional sentences can be considered to be a set of all interpretations which satisfy those sentences, that is, a set of models and a belief can be regarded as an element of the set.

In this paper, contradiction does not mean logical contradiction; *knowledge* must not contain any contradiction. However, a combination of *knowledge* and *belief* may cause a contradiction.

The formalization of *knowledge* and *belief* is as follows.

Let U be a finite set of propositional symbols. The set of *well formed formulas* (this set is denoted by Wff) is defined below:

(1) $U \subseteq Wff$,
(2) $\alpha \in Wff$ implies $\neg\alpha \in Wff$,
(3) $\alpha,\beta \in Wff$ implies $\alpha \supset \beta \in Wff$,
(4) Nothing else is a *well formed formula.*

The following abbreviations are used:

(1) $\alpha \vee \beta = \neg\alpha \supset \beta$,
(2) $\alpha \wedge \beta = \neg(\alpha \supset \neg\beta)$.

An *interpretation* is defined as a subset of U. Every propositional symbol in the subset is supposed to be assigned to 1 or *true*. A set of all *interpretations* (that is 2^U) is denoted as Ω. The relation "$\models$" $\subseteq \Omega \times Wff$ is defined as follows:

Let $I \in \Omega$
(1) If $\alpha \in U$ then $I \models \alpha$ iff $\alpha \in I$,
(2) If $\alpha = \neg\beta$ then $I \models \alpha$ iff not $I \models \beta$,
(3) If $\alpha = \beta \supset \gamma$ then $I \models \alpha$ iff not $I \models \beta$ or $I \models \gamma$.

The following notation is also used:

Let $I \in \Omega$ and $\Gamma \subseteq Wff$, $I \models \Gamma$ iff $I \models \alpha$ for all $\alpha \in \Gamma$,
Let $I \in \Omega$ and $\alpha \in Wff$, $I \not\models \alpha$ iff not $I \models \alpha$.

Reasoning from incomplete knowledge with belief is now defined.

True reasoning from *knowledge* is first defined. The relation "$\models$" $\subseteq (\Omega \times 2^{Wff}) \times Wff$ is defined as follows:

$(I,\Gamma) \models \alpha$ iff $I \models \Gamma$ and for all $J \in \Omega$, if $J \models \Gamma$ then $J \models \alpha$.

This definition is similar to an ordinary *entailment* definition.

Temporary reasoning from the belief is defined. The relation "$\approx$" over $(\Omega \times 2^{Wff}) \times Wff$ is defined as follows:

$(I,\Gamma) \approx \alpha$ iff $I \models \Gamma$ and $I \models \alpha$ and there exists $J \in \Omega$ such that $J \models \Gamma$ and $J \not\models \alpha$.

This definition expresses that even if *knowledge* Γ does not entail query α, we can draw a conclusion from *belief I*.

3. A Criterion of Belief Revision

This section defines belief revision criterion based on *minimal change*. When new knowledge is added, it may cause contradiction with a belief. If it happens, we must change the belief to keep consistency. However, we must decide how we should change it. Although we must change the part of the belief which leads to contradiction, we do not need to change an irrelevant part, that is, a part whose change does not help to restore consistency. This situation is defined as *minimal change* in terms of a set difference between a current belief and a new belief.

A *belief revision relation* (BR) is a relation over $\Omega \times 2^{Wff} \times Wff \times \Omega$ defined below:

$(I, \Gamma, \alpha, J) \in BR$ iff
(1) $I \models \Gamma$,
(2) $J \models \Gamma \cup \{\alpha\}$,
(3) For all $K \in \Omega$, if $K \models \Gamma \cup \{\alpha\}$ and $I \cap J \subseteq K \subseteq I \cup J$ then $J = K$.

Informally, J is a model of new *knowledge* and an interpretation that is minimally changed from I.

We write $(I, \Gamma) \stackrel{\alpha}{\Longrightarrow} (J, \Gamma \cup \{\alpha\})$ iff $(I, \Gamma, \alpha, J) \in BR$.

4. Characteristics of Belief Revision

This section outlines the characteristics of *minimal change belief revision*.

Theorem 1: *Let* $(I, \Gamma) \stackrel{\alpha}{\Longrightarrow} (J, \Gamma \cup \{\alpha\})$ *and* $I \models \alpha$*, then* $J = I$.

Proof: Suppose that $I \neq J$, it is shown that J does not satisfy condition (3) of the definition. Take I as K of (3). $I \models \Gamma \cup \{\alpha\}$ and $I \cap J \subseteq I \subseteq I \cup J$, but $I \neq J$.

This theorem states that if the current belief is consistent with added information, belief revision is not required.

Theorem 2: *Let* K_1 *be any subset of* $J - I$ *and* K_2 *be any subset of* $I - J$*. If* $(I, \Gamma) \stackrel{\alpha}{\Longrightarrow} (J, \Gamma \cup \{\alpha\})$ *and* $J \neq (J - K_1) \cup K_2$ *then* $(J - K_1) \cup K_2 \not\models \Gamma \cup \{\alpha\}$

Proof: Suppose that there exists $K_1 \subseteq J - I$ and $K_2 \subseteq I - J$ and $J \neq (J - K_1) \cup K_2$ and $(J - K_1) \cup K_2 \models \Gamma \cup \{\alpha\}$. It is shown that J does not satisfy condition (3) of the definition. Clearly, $I \cap J = I \cap (J - K_1)$ since $K_1 \cap I = \phi$. Let $K = (J - K_1) \cup K_2$, then $K \models \Gamma \cup \{\alpha\}$ and $I \cap J = I \cap (J - K_1) \subseteq I \cap ((J - K_1) \cup K_2) = I \cap K \subseteq K \subseteq I \cup K = I \cup ((J - K_1) \cup K_2) = (I \cup K_2) \cup (J - K_1) = I \cup (J - K_1) \subseteq$

$I \cup J$, but $K \neq J$.

This theorem states that all changes of interpretation in the new belief are minimally necessary for maintaining consistency.

5. System Overview

This section describes the system based on the above semantics.

Initially, the user gives *knowledge* (which is a subset of 2^{Wff}) and a model of the *knowledge* as a *belief* (which is a subset of Ω). Then the user asks a *query* (which is a wff) or adds new *knowledge* (which is also a wff) to the system.

When the user asks *query* α to the system, the system may answer in four different ways (let the current *knowledge* be Γ and the current *belief* be I).

(1) If the current *knowledge* entails the *query* (that is, $(I, \Gamma) \models \alpha$), the system answers "yes".
(2) If the current *knowledge* entails the negation of the *query* (that is, $(I, \Gamma) \models \neg \alpha$), the system answers "no".
(3) If neither case (1) nor case (2) holds but using the *belief* makes the *query* true (that is, $(I, \Gamma) \approx \alpha$), then the system answers "I believe".
(4) If neither case (1) nor case (2) holds but using the *belief* makes the *query* false (that is, $(I, \Gamma) \approx \neg \alpha$), then the system answers "I believe not".

When new *knowledge* α is asserted, the system will react the following way.

(1) If the new *knowledge* is contradictory to the current *knowledge* (that is, $(I, \Gamma) \models \neg \alpha$), it will be rejected.
(2) If the new *knowledge* is consistent with the current *belief* (that is, $I \models \alpha$), the current *belief* need not be changed by Theorem 1.
(3) If the new *knowledge* is not consistent with the current *belief* (that is, $I \not\models \alpha$), the system revises the current *belief* and changes it to a minimally different *belief* which is consistent with both the current *knowledge* and the new *knowledge*. (The system finds a new *belief* J such that $(I, \Gamma) \stackrel{\alpha}{\Longrightarrow} (J, \Gamma \cup \{\alpha\})$.)

6. Example

This section shows an example of belief revision, taken from [Doyle79]. Consider the time and place of a meeting. A meeting will be held at either 10 a.m. or 11 a.m. and in either room 813 or 801. The set of propositions is;

$$U = \{MeetingAt10, MeetingAt11, MeetingIn813, MeetingIn801\}$$

First, we give an initial belief, I_0, and initial knowledge, Γ_0.

$I_0 = \{MeetingAt10, MeetingIn813\}$
$\Gamma_0 = \phi$

I_0 expresses that we believe a meeting will be at 10 a.m. in room 813 (and not at 11 a.m. and not in room 801). Currently, there is no knowledge, so the system cannot give a true reply, but it can answer a believed result. We ask "Will the meeting be held at neither 10 a.m. nor 11 a.m.?" by giving the following query to the system.

$$\chi_0 = \neg\ MeetingAt10 \wedge \neg\ MeetingAt11$$

Since $(I_0, \Gamma_0) \approx\neg\ \chi_0$, the system will reply "I believe not".

And we give knowledge. Since the meeting will be held at either 10 a.m. or 11 a.m., we give knowledge α_0 to the system.

$\alpha_0 = MeetingAt10 \vee MeetingAt11$
$\Gamma_1 = \{\alpha_0\}$

Since $I_0 \models \alpha_0$, the current belief is not changed. When we make the same query of χ_0 to the system, this time it will reply "no" as $(I_0, \Gamma_1) \models \neg\ \chi_0$. We assert that the meeting will be held in either room 813 or 801.

$\alpha_1 = MeetingIn813 \vee MeetingIn801$
$\Gamma_2 = \{\alpha_0, \alpha_1\}$

Since $I_0 \models \alpha_1$, the current belief is not changed. Then we find that we cannot use the room 813 at 10 a.m. We give the following knowledge α_2 to the system.

$\alpha_2 = \neg\ (MeetingAt10 \wedge MeetingIn813)$
$\Gamma_3 = \{\alpha_0, \alpha_1, \alpha_2\}$

Since $I_0 \not\models \alpha_2$, belief revision occurs and the system tries to find J such that $(I_0, \Gamma_2) \stackrel{\alpha}{\Longrightarrow} (J, \Gamma_3)$. There are two minimal changed beliefs.

$J_0 = \{MeetingAt11, MeetingIn813\}$
$J_1 = \{MeetingAt10, MeetingIn801\}$

(Note that $J_2 = \{MeetingAt11, MeetingIn801\} \models \Gamma_3$, but that J_2 is not a minimal changed model since $J_2 \cap I_0 \subseteq J_0$ (or J_1) $\subseteq J_2 \cup I_0$.)

The system selects one of them. Suppose J_0 is selected. Then it is found that the meeting cannot be held at 11 a.m. We give that fact to the system.

$\alpha_3 = \neg\ MeetingAt11$
$\Gamma_4 = \{\alpha_0, \alpha_1, \alpha_2, \alpha_3\}$

Since $J_0 \not\models \alpha_3$, belief revision occurs again and the system finds the following new belief, I_1.

$I_1 = \{MeetingAt10, MeetingIn801\}$

7. Related Research

7.1 Doyle's TMS

The last section shows that our system seems to perform the same tasks as Doyle's TMS [Doyle79] in the propositional case, but more simply and clearly. Since his TMS concentrated on the efficient management of finding the source of contradiction, the data structure became complex and as a result, formal analysis is very difficult.

[Doyle83] tries to formalize TMS in terms of *nodes* and uses a similar notion of closeness of the state when *conditional proof* occurs. However, Doyle's formalization is not to provide the simple semantics of his system, but to prove the correctness of his mechanism of truth maintenance. Contrary to his approach, we first try to formalize belief revision in a logical framework without thinking about efficient implementation since we think that it is important to give a simple specification of belief revision to analyze. However, his TMS may be useful for efficient implementation of our system.

7.2 MBR

The Multiple Belief-Space Reasoner (MBR) [Martins84] is similar to our approach because it is based on the logic called the SWM system. The SWM system is logic which can manipulate truth maintenance in itself. The objects in the system are called *supported wffs*, and consist of a wff and an associated triple (OT, OS, and RS). OT is an origin tag which expresses the type of origin of a wff, OS is an origin set which is a source of the wff, and RS is a restriction set which causes contradiction if the wff and any element of RS are in the current belief at the same time. It provides inference rules which manage OT, OS, and RS in order to find and avoid contradiction. Like Doyle's formalism, its inference rules are the formalization of an abstract implementation.

7.3 ATMS

The ATMS [de Kleer86] concentrates on efficient management of multiple belief spaces by avoiding contradiction when assumptions are combined. [Reiter87] provides a formalism of hypothetical reasoning of ATMS. Reiter formalizes that the ATMS answers the simplest combination of hypotheses (called *minimal support*) of a query. Although it is not a formalism of belief revision, the semantics of hypothetical reasoning are very clear and easy to analyze. Our formalism also gives clear and simple semantics for belief revision.

7.4 Nonmonotonic Reasoning

Nonmonotonic reasoning[McCarthy80, McDermott80, Reiter80] performs a kind of belief revision since a result which could be derived previously is no longer derived if added information is inconsistent with the result. Of those formalisms, [Reiter80] is closely related to ours. [Reiter80] defines a belief, called an *extension*, as a set of consistent formulas with axioms and *default rules*, and provides a mechanism to decide whether belief revision is needed or not when new information is added. However, it does not provide any criteria for belief revision.

8. Conclusion

This paper presented a clear and simple system which performs reasoning from incomplete knowledge with beliefs, and a criterion of belief revision. It defined knowledge as a set of propositions and a belief as a model of that set in order to answer every query from the user by using the belief. However, the belief is not always true and if it is found to be false, that is, if it turns out to be inconsistent with new knowledge, it must be revised. A criterion called *minimal change* was presented in terms of a set difference between a current belief and a new belief. *Minimal change* formalizes that we change our new belief to keep it as close to our previous belief as possible.

References

[de Kleer86] de Kleer, J., "An Assumption-based TMS", *Artificial Intelligence*, Vol. 28, No.2, 1986

[Doyle79] Doyle, J., "A Truth Maintenance System", *Artificial Intelligence*, Vol. 12, No.3, 1979

[Doyle83] Doyle, J., "The Ins and Outs of Reason Maintenance", *Proc. of IJCAI-83*, 1983

[Martins84] Martins, J. and Shapiro, S., "A Model for Belief Revision", *Proc. of Non-monotonic Reasoning Workshop*, 1984

[Martins86] Martins, J. and Shapiro, S., "Theoretical Foundation for Belief Revision", *Proc. of Theoretical Aspects of Reasoning about Knowledge*, 1986

[McCarthy80] McCarthy, J., "Circumscription - a Form of Non-monotonic Reasoning", *Artificial Intelligence*, Vol. 13, 1980

[McDermott80] McDermott, D. and Doyle, J., "Non-monotonic Logic I", *Artificial Intelligence*, Vol. 13, 1980

[Reiter80] Reiter, R., "A Logic for Default Reasoning", *Artificial Intelligence*, Vol. 13, 1980

[Reiter87] Reiter, R. and de Kleer, J., "Foundations of Assumption-based Truth Maintenance System: Preliminary Report", *Proc. of AAAI-87*, 1987