Abductive Framework for Nonmonotonic Theory Change

Katsumi Inoue
Department of Information and Computer Sciences
Toyohashi University of Technology
Tempaku-cho, Toyohashi 441, Japan
inoue@tutics.tut.ac.jp

Chiaki Sakama
Department of Computer and Communication Science
Wakayama University
930 Sakaedani, Wakayama 640, Japan
sakama@center. wakayama-u .ac.jp

Abstract

This paper proposes a method of nonmonotonic theory change. We first introduce a new form of abduction that can account for observations in nonmonotonic situation. Then we provide a framework of autoepistemic update, which describes nonmonotonic theory change through the extended abductive framework. The proposed update semantics is fairly general and provides a unified framework for various update semantics such as first-order update, view update of databases, and contradiction removal of nonmonotonic theories.

1 Introduction

A lot of theories for belief change have been proposed in AI and related fields. At abstract and philosophical levels, the belief dynamics have been studied as *rationality postulates* to be satisfied by belief sets (e.g., [Alchourfon *et* a/., 1985; Katsuno and Mendelzon, 1991b]). In the field of AI and databases, various researchers have proposed *revision, contraction,* and *update* methods of data and knowledge bases. On the AI side, revision and update methods mainly cope with knowledge bases which consist of first-order theories. According to [Winslett, 1990], those methods are classified into *formula-based* and *model-based* approaches. In the formula-based update such as [Fagin *et al.,* 1983], the units of change are formulas, and the syntax of formulas in a theory influences the result of update. In the model-based update, on the other hand, update does not care about formulas in a theory but cares about changes of models during update. In both approaches, however, the underlying language for describing beliefs is a monotonic (mainly propositional) logic. In fact, not much is known about *update of nonmonotonic theories.* Note that this fact should not be confused with the well-known fact that the process of belief change itself is nonmonotonic even when our beliefs are represented in monotonic logic [Makinson and Gardenfors, 1989].

Using nonmonotonic logics, one expects that some previously derived formulas are automatically retracted when our belief set changes. However, the present nonmonotonic formalisms are not strong enough to revise even a simple belief set. For instance, let us consider an example (due to [Konolige, 1992]) of autoepistemic logic, where an agent's belief is given by the theory K_0-

$$K_0 = \{ \neg Bp \supset q \}.$$

The formula in K_0 can be read as: if she does not believe that "the repairman has arrived (p)" then "the copier must be OK ((q)". Suppose further that she then found that "the copier is broken (-q)". Then her beliefs are revised as

$$K_1 = K_0 \cup \{ \neg q \} = \{ \neg q, \neg Bp \supset q \}.$$

The theory K_1 now does not have any stable expansion, because while Bp is derivable p is not. We thus need a revision method for nonmonotonic theories that can retract previously derived formulas automatically.

In the context of databases, on the other side, update of deductive databases is usually captured as the *view update* problem [Abiteboul, 1988]. Namely, in a deductive database, update on virtual relations in an *intensional database* has to be translated into update on real facts in an *extensional database.* It is also known that database update is closely related to *abduction* in AI. Kakas and Mancarella [1990] present that view update in deductive databases is realized by an abductive procedure of logic programming by considering update requests as observations and extensional relations as abductive hypotheses. This close relationship between abduction and update, however, need not be limited within the area of deductive databases. We consider that abduction can play a fundamental role in a wide class of AI and database problems. That is, abductive methods would contribute to better understanding of various belief change semantics as well as better implementation of them.

In this paper, we characterize update of nonmonotonic theories through abduction. For this purpose, we first extend the abductive framework to autoepistemic theories. The notions of *negative explanations* and *anti-explanations* are introduced to account for observations in nonmonotonic setting. Then we define *autoepistemic update* through the extended abductive framework. It is shown that autoepistemic update can provide a uniform framework for various update semantics. In particular, update of first-order theories, view update of deductive databases, and contradiction removal of nonmonotonic theories are expressed as special cases of the new update semantics.

The rest of this paper is organized as follows. The next section reviews autoepistemic logic. Section 3 defines a new abductive framework for autoepistemic logic. The abductive framework is applied to formalize an update semantics for autoepistemic theories in Section 4. Various forms of update semantics are expressed in terms of autoepistemic update. Section 5 discusses related work, and Section 6 concludes the paper.

2 Autoepistemic Theory

We briefly review autoepistemic logic by Moore [1985]. Autoepistemic logic is obtained by extending a first-order language $\mathcal{L}_0$ with the modal operator B. A formula in autoepistemic logic is called objective if it does not contain the modal operator B; otherwise it is subjective. Intuitively, the formula BF is read as "F is believed". By an autoepistemic theory, or simply a theory, we mean a set of formulas in autoepistemic logic. In this paper, we allow open variables, and each formula with variables stands for the set of its ground instances. This means that an autoepistemic theory is essentially equivalent to a set of countably many propositional formulas in autoepistemic logic.

A theory is stable if it is closed under the logical and introspective consequences. Namely, a stable set T satisfies the conditions: (i) $T = cons(T)$, where cons(T) denotes the set of logical consequences (in the sense of classical first-order logic) of T; (ii) if $\varphi \in T$ then $B\varphi \in T$, and (iii) if $\varphi \notin T$ then $\neg B\varphi \in T$. The meaning of each autoepistemic theory is usually characterized by the following stable set that is expanded from the theory: Given an autoepistemic theory K, a set T is a stable expansion of K iff it satisfies that

$$T = cons(K \cup \{B\varphi \mid \varphi \in T\} \cup \{\neg B\varphi \mid \varphi \notin T\}).$$

Note that an autoepistemic theory may have none, one, or multiple stable expansions. We say that an autoepistemic theory K is consistent if it has a consistent stable expansion; otherwise K is inconsistent

It is well known that for each set Σ of objective formulas, there is a unique stable set $E(\Sigma)$ containing Σ such that the objective formulas in $E(\Sigma)$ are exactly the same as those in $cons(\Sigma)$. Moreover, if a theory K contains only objective formulas, then $E(K)$ is a unique stable expansion of A" [Moore, 1985].

By $K \models F$, we mean that every stable expansion of an autoepistemic theory K contains a formula F. For example, $\{\neg Bp \supset q\} \models q$ and $\{\neg Bp \supset q, \ \neg Bq \supset p\} \not\models q$. Thus $K \not\models F$ denotes that there is a stable expansion T of K such that T does not include F. This skeptical reading of the entailment relation in autoepistemic logic generalizes the meaning of the classical entailment relation.[1] Namely, for a first-order theory Σ and a first-order formula F, it holds that $\Sigma \models F$ iff $F \in E(\Sigma)$.

[1] We can give an alternative, credulous meaning to K = F: that is, there is a stable expansion of K containing F. Note that in this weaker reading, again the notion is a generalization of the first-order entailment relation since any first-order theory has the unique stable expansion.

Autoepistemic logic and the notion of stable expansions have a close relationship with the answer set semantics [Gelfond and Lifschitz, 1991] for logic programming with negation as failure and classical negation (see [Lifschitz and Schwarz, 1993; Inoue and Sakama, 1994], for instance). Hence, the theory of belief update developed in this paper can directly be applied to the update problem for such extended logic programs.

3 New Form of Abduction

Abduction is one of the three fundamental modes of reasoning characterized by C. S. Peirce, the others being deduction and induction. The most popular formalization of abduction in AI defines an explanation as a set of hypotheses which, together with the background theory, logically entails the given observations. A traditional, logical framework of abduction is defined as follows. Let Σ (background theory) and T (hypotheses) be two sets of first-order formulas. Given a formula F (observation), a set E of ground instances of elements from T is an explanation of F with respect to $\langle \Sigma, \Gamma \rangle$ if

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

An explanation E is minimal if no $E' \subset E$ is an explanation of F. Suppose, for example, that we are given the background theory Σ_1 and the hypotheses T1 as

$$\begin{aligned} \Sigma_1 = \{ \quad & bird(x) \wedge \neg ab(x) \supset flies(x), \\ & bird(x) \wedge ab(x) \supset \neg flies(x), \\ & bird(tweety) \quad \}, \\ \Gamma_1 = \{ \quad & \neg ab(x), \quad ab(x) \quad \}. \end{aligned}$$

Here, the hypothesis -ab(x) means that for any ground term t, -ab(t) can be hypothesized. In other words, a hypothesis containing variables is shorthand for the set of its ground instances. In this case, a minimal explanation of the observation flies(tweety) is $\{\neg ab(tweety)\}$ As a result of assimilating the observation flies(tweety), our background theory is changed as

$$\Sigma_2 = \Sigma_1 \cup \{ \neg ab(tweety) \}.$$

Suppose we later find that tweety losts his flying ability for some reasons (e.g., injured, fatted, etc). In this case, -flies(tweety) should be explained by assuming ab(tweety) instead of -ab(tweety). Retracting the previous assumption -ab(tweety) is vital, since flies(tweety) should not be explained any more. To formalize such a situation, we extend the above abductive framework in the following three respects:

1. The background theory Σ and the candidate hypotheses T can be autoepistemic theories. Thus, the belief operator B may appear in Σ or Γ.
2. Hypotheses can not only be added to the theory Σ but also be discarded from Σ to explain observations. When $\Sigma \setminus E \models F$ for some set E of hypotheses, we call E a negative explanation of F. An ordinary explanation E such that $\Sigma \cup E \models F$ is now called a positive explanation.

3. The notion of explanation is extended to *unexplain* observations. When $\Sigma \cup E \not\models F$ (resp. $\Sigma \setminus E \not\models F$) for some set E of hypotheses, we call E a positive (resp. negative) *anti-explanation.*

The first extension comes from a demand of *commonsense reasoning* that we want to treat a nonmonotonic theory as a background belief theory. An example of such an extension in the literature can be seen in Inoue's abductive framework [Inoue, 1994] where both Σ and Γ are extended logic programs. Then, it is natural to consider the second extension. This is because the background theory is nonmonotonic, so that contraction of some hypotheses may "revive" the previously derived conclusions. Although it is very natural, as far as the authors know, the notion of negative explanation has never been proposed in the literature. The third extension is useful to talk about unexplained formulas that had been explained before hypotheses changed positively or negatively.

Positive and negative (anti-)explanations are often combined to (un)explain observations. We thus formally define an explanation for our abductive framework by means of a pair of positive and negative explanations.

Definition 3.1 An *abductive framework* is a pair $\langle K, \Gamma \rangle$ in which both K and Γ are autoepistemic theories. Let F be a formula. A pair (I, O) is an *explanation* (resp. *anti-explanation*) of F (with respect to $\langle K, \Gamma \rangle$) if

1. $(K \cup I) \setminus O \models F$ (resp. $(K \cup I) \setminus O \not\models F$),
2. $(K \cup I) \setminus O$ is consistent, and
3. both I and O consist of instances of elements from Γ.

An (anti-)explanation (I, O) of F is *minimal* if for any explanation (I', O') of F, $I' \subseteq I$ and $O' \subseteq O$ imply that $I' = I$ and $O' = O$.

Intuitively, a minimal (anti-)explanation (I, O) offers a tradeoff between *minimal reduction* O of the background theory K and *minimal augmentation* with the newly added hypotheses I from Γ. In this sense, each minimal (anti-)explanation accomplishes a *minimal change* of the theory K with respect to Γ. Obviously, this definition of (minimal) explanations reduces to the traditional one when K is a first-order theory and O for each explanation (I, O) is empty. For the bird example above, a minimal explanation of $flies(tweety)$ with respect to $\langle \Sigma_1, \Gamma_1 \rangle$ is $(\{\neg ab(tweety)\}, \emptyset)$. Also, a minimal anti-explanation of $flies(tweety)$ with respect to $\langle \Sigma_2, \Gamma_1 \rangle$ is $(\emptyset, \{\neg ab(tweety)\})$, and a minimal explanation of $\neg flies(tweety)$ with respect to $\langle \Sigma_2, \Gamma_1 \rangle$ is $(\{ab(tweety)\}, \{\neg ab(tweety)\})$. Notice that an anti-explanation of F is not an explanation of $\neg F$, and that anti-explanations cannot be represented by explanations in general.

Example 3.2 Let us consider the autoepistemic theories introduced in Section 1:

$$K_0 = \{ \neg Bp \supset q \}, \quad K_1 = K_0 \cup \{\neg q\}$$

with the hypotheses $\Gamma_0 = \{p\}$. The (minimal) anti-explanation of q with respect to (K_0, Γ_0) is $(\{p\}, \emptyset)$, which is also the (minimal) explanation of $\neg q$ with respect to (K_1, Γ_0).

In the next section, we define an update semantics of autoepistemic theories in terms of the extended abductive framework.

4 Autoepistemic Update

A formula in autoepistemic logic is called a *dynamic belief* if it is subject to change in an autoepistemic theory. We denote as Γ the set of all dynamic beliefs in the language. In this setting, an autoepistemic theory K is divided into two parts based on this meta-theoretical partition of beliefs: (i) the *invariant beliefs* K_I and (ii) the *dynamic beliefs* K_D. Namely, K can be written as

$$K = K_I \cup K_D, \text{ where } K_D = K \cap \Gamma \text{ and } K_I = K \setminus K_D.$$

The notion of such a partition (with possibly more than two levels) of beliefs is commonly used for update semantics in the literature: for example, *tagged sentences* [Fagin *et al.*, 1983], the notion of *(un)protected formulas* [Winslett, 1988], *integrity constraints* [Katsuno and Mendelzon, 1991a], and *priority classes* [Nebel, 1991]. Thus update to a theory should be translated into update on dynamic beliefs that an agent can actually change.

We define *update* as a function u between autoepistemic theories.[2] Given an *initial theory* K, we call $K' = u(K)$ an *updated theory*. Each update considered in this paper is either an *insertion* or a *deletion* of a formula in autoepistemic logic. Now, autoepistemic update is defined as follows.

Definition 4.1 Given a theory K, the *autoepistemic update* is defined in terms of the abductive framework as follows.

1. If update is an *insertion* of a formula F into K, then $u(K) = (K \cup I) \setminus O$ where (I, O) is a minimal explanation of F with respect to $\langle K, \Gamma \rangle$.
2. If update is a *deletion* of a formula F from K, then $u(K) = (K \cup I) \setminus O$ where (I, O) is a minimal anti-explanation of F with respect to $\langle K, \Gamma \rangle$.

If there is no such (anti-)explanation, update is *impossible*.

Example 4.2 (cont. from Example 3.2) The deletion of q from the theory $K_0 = \{\neg Bp \supset q\}$ whose dynamic beliefs are $\Gamma_0 = \{p\}$ is accomplished by the anti-explanation $(\{p\}, \emptyset)$ of q with respect to (K_0, Γ_0).

Example 4.3 Consider the theory

$$K_2 = \{ q \wedge r \supset p, \; Bq \vee r \} \text{ with } \Gamma_2 = \{q\}.$$

In this theory, the insertion of p is impossible. In fact, there is no pair (I, O) that accomplishes the insertion.

The proposed autoepistemic update is general enough to provide a unified framework for various update semantics. In the following, we present the relationships between autoepistemic update and those other semantics.

[2] According to [del Val and Shoham, 1994], *revision* is considered as update of mental states. With this regard, update of autoepistemic theories is also considered as revision of belief states. So we do not distinguish revision and update unlike [Katsuno and Mendelzon, 1991b], and use the terms interchangeably.

4.1 First-Order Update

First-order theory update is a special case of autoepistemic update when K and F are first-order theories in Definition 4.1. Some work have been done for updating first-order theories. Among them, Fagin *et al.*'s [1983] semantics for updating first-order theories is one of the best-known frameworks in the field. In this subsection, we address a method of realizing their update semantics in our abductive framework.

Definition 4.4 Let Σ be a first-order theory and σ a formula. A theory $u_F(\Sigma)$ accomplishes a *FUV insertion* of σ into Σ if

1. $\{\sigma\} \subseteq u_F(\Sigma) \subseteq \Sigma \cup \{\sigma\}$,
2. $u_F(\Sigma)$ is consistent, and
3. there is no set $u'_F(\Sigma)$ satisfying the above two and $u_F(\Sigma) \subset u'_F(\Sigma)$.

Note that in a FUV insertion an inserted formula σ is always included in the updated theory $u_F(\Sigma)$,[3] and any formula in Σ is considered as a retractable formula. Thus the updated theory is defined as a union of the inserted formula σ and a maximal subset of the original formulas Σ that is consistent with σ. This situation is realized in our abductive framework by inserting σ to the original theory and changing any formula other than σ minimally to make the updated theory consistent.

Theorem 4.5 *Given a first-order theory Σ and a formula σ, $u_F(\Sigma)$ accomplishes a FUV insertion of σ into Σ iff $u_F(\Sigma) = (\Sigma \cup \{\sigma\}) \setminus O$ where $(\emptyset, O)$ is a minimal explanation of σ with respect to $\langle \Sigma \cup \{\sigma\}, \Sigma \setminus \{\sigma\} \rangle$.*

Proof: Suppose that $u_F(\Sigma)$ accomplishes a FUV insertion of σ into Σ. $u_F(\Sigma)$ can be written as $(\Sigma \cup \{\sigma\}) \setminus O$, where $\sigma \notin O$, because $\sigma \in u_F(\Sigma)$. Then, $(\Sigma \cup \{\sigma\}) \setminus O \models \sigma$ and $(\Sigma \cup \{\sigma\}) \setminus O$ is consistent. Therefore, O is a negative explanation of σ with respect to $\langle \Sigma \cup \{\sigma\}, \Sigma \setminus \{\sigma\} \rangle$. Suppose to the contrary that $(\emptyset, O)$ is not a minimal explanation of σ. Then, there exists an explanation (I', O') of σ such that $I' \subseteq \emptyset$ and $O' \subseteq O$, but $I' \neq \emptyset$ or $O' \neq O$. The last condition implies that $O' \subset O$. In this case, $(\Sigma \cup \{\sigma\}) \setminus O' \models \sigma$ because of the monotonicity and $O' \subset O$. Likewise, $(\Sigma \cup \{\sigma\}) \setminus O'$ is consistent. Then, $u'_F(\Sigma) = (\Sigma \cup \{\sigma\}) \setminus O'$ satisfies the condition 1 and 2 of Definition 4.4, and $u_F(\Sigma) \subset u'_F(\Sigma)$ holds. This contradicts the fact that $u_F(\Sigma)$ accomplishes a FUV insertion of σ into Σ. Hence, $(\emptyset, O)$ is a minimal explanation of σ.

Conversely, suppose that $(\emptyset, O)$ is a minimal explanation of σ with respect to $\langle \Sigma \cup \{\sigma\}, \Sigma \setminus \{\sigma\} \rangle$. Obviously, $\{\sigma\} \subseteq [(\Sigma \cup \{\sigma\}) \setminus O] \subseteq \Sigma \cup \{\sigma\}$, and $(\Sigma \cup \{\sigma\}) \setminus O$ is consistent. Since $(\emptyset, O)$ is a minimal explanation of σ, no $(\emptyset, O')$ such that $O' \subset O$ is an explanation of σ. Therefore, $u_F(\Sigma) = (\Sigma \cup \{\sigma\}) \setminus O$ accomplishes a FUV insertion of σ into Σ. □

Fagin *et al.*'s deletion is not defined as the counterpart of their insertion, because formulas to be deleted may be logical consequences of the theory.

[3] As discussed in [Winslett, 1990, Section 2.1.2], adding an insertion formula σ to Σ is not always desirable. With this regard, our update method by Definition 4.1 can accomplish an insertion in a more flexible manner.

Definition 4.6 Let Σ be a first-order theory and σ a non-tautological formula. A theory $u_F(\Sigma)$ accomplishes a *FUV deletion* of σ from Σ if

1. $u_F(\Sigma) \subseteq \Sigma$,
2. $u_F(\Sigma) \not\models \sigma$, and
3. there is no set $u'_F(\Sigma)$ satisfying the above two and $u_F(\Sigma) \subset u'_F(\Sigma)$.

In the above definition an updated theory is defined as a maximal subset S of the theory Σ such that σ does not follow from S. This situation is also expressed in our abductive framework by minimizing a set of discarded hypotheses from Σ to unexplain σ.

Theorem 4.7 *A theory $u_F(\Sigma)$ accomplishes a FUV deletion of a formula σ from the theory Σ iff $u_F(\Sigma) = \Sigma \setminus O$ where $(\emptyset, O)$ is a minimal anti-explanation of σ with respect to $\langle \Sigma, \Sigma \rangle$.*

Proof: Suppose that $u_F(\Sigma)$ accomplishes a FUV deletion of σ from Σ. $u_F(\Sigma)$ can be written as $\Sigma \setminus O$ where $O \subseteq \Sigma$. By definition, $\Sigma \setminus O \not\models \sigma$, and hence $\Sigma \setminus O$ is consistent. Therefore, O is a negative anti-explanation of σ with respect to $\langle \Sigma, \Sigma \rangle$. Suppose to the contrary that $(\emptyset, O)$ is not a minimal anti-explanation of σ. Then, there exists an anti-explanation (I', O') of σ such that $O' \subset O$. Put $u'_F(\Sigma) = \Sigma \setminus O'$. By definition, $u'_F(\Sigma) \not\models \sigma$. Obviously, $u'_F(\Sigma) \subseteq \Sigma$, but $u_F(\Sigma) \subset u'_F(\Sigma)$. This contradicts the fact that $u_F(\Sigma)$ accomplishes a FUV deletion of σ from Σ.

Conversely, suppose that $(\emptyset, O)$ is a minimal anti-explanation of σ with respect to $\langle \Sigma, \Sigma \rangle$. Obviously, $(\Sigma \setminus O) \subseteq \Sigma$. By definition, $\Sigma \setminus O \not\models \sigma$. Since $(\emptyset, O)$ is a minimal anti-explanation of σ, no $(\emptyset, O')$ such that $O' \subset O$ is an anti-explanation of σ. Therefore, $u_F(\Sigma) = \Sigma \setminus O$ accomplishes a FUV deletion of σ from Σ. □

4.2 View Update of Databases

In the context of databases, update is usually captured as the *view update* problem, in which update on intensional facts is translated into update on extensional facts in a database. Here, we characterize view update in our abductive framework.

A *database* D is a finite set of rules of the form:

$$A \leftarrow A_1, \ldots, A_m, not\, A_{m+1}, \ldots, not\, A_n \quad (n \geq m \geq 0)$$

where A and A_i's are atoms (A is possibly empty) and *not* denotes the negation-as-failure operator in logic programming. When $n = 0$, the rule is also written just as A by omiting $\leftarrow$. Each rule of the above form is regarded as the autoepistemic formula

$$A_1 \wedge \ldots \wedge A_m \wedge \neg BA_{m+1} \wedge \ldots \wedge \neg BA_n \supset A,$$

where $A = false$ when A is empty in the rule.

Given a database D, $\mathcal{EB}$ is the pre-specified set of atoms from the language of D called the *extensional base*. Then view update of a database is defined as follows.

Definition 4.8 Let D be a database and A a ground atom. An *insertion* of A into D (resp. *deletion* of A from D) is accomplished by an updated database $u(D) = (D \cup I) \setminus O$, where $I, O \subseteq \mathcal{EB}$, if

1. $u(D) \models A$ (resp. $u(D) \not\models A$),
2. $u(D)$ is consistent,
3. there is no $u'(D) = (D \cup I') \setminus O'$, where $I', O' \subseteq \mathcal{EB}$, satisfying the above two and $\Delta_{D,u'(D)} \subset \Delta_{D,u(D)}$, where $\Delta_{D_1,D_2} = (D_1 \setminus D_2) \cup (D_2 \setminus D_1)$.

Note that the condition 3 in the above definition says that update of D is performed minimally on $\mathcal{EB}$. The view update problem of deductive databases is characterized by abduction as follows.

Lemma 4.9 *Let $u(D) = (D \cup I) \setminus O$ be an updated database that accomplishes an insertion/deletion of an atom A into/from D as in Definition 4.8. Then,*

1. $I \cap O = \emptyset$,
2. $I \cap D = \emptyset$, *and*
3. $O \subseteq D$.

Theorem 4.10 *Given a database D and a ground atom A, a database $u(D)$ accomplishes an insertion (resp. deletion) of A into/from D iff A has a minimal explanation (resp. anti-explanation) (I, O) of A with respect to $\langle D, \mathcal{EB} \rangle$.*

Proof: We consider the insertion of A since the case of the deletion is similar. Suppose that $u(D) = (D \cup I) \setminus O$, where $I, O \subseteq \mathcal{EB}$, accomplishes an insertion of A into D. By definition, (i) $(D \cup I) \setminus O \models A$, (ii) $(D \cup I) \setminus O$ is consistent, and (iii) no other (I', O') satisfies (i), (ii) and

$$[D \setminus \{(D \cup I') \setminus O'\}] \cup [\{(D \cup I') \setminus O'\} \setminus D] \subset [D \setminus \{(D \cup I) \setminus O\}] \cup [\{(D \cup I) \setminus O\} \setminus D].$$

By (i) and (ii), (I, O) is an explanation of A with respect to $\langle D, \mathcal{EB} \rangle$. The last relation in the condition (iii) is equivalent to

$$(D \cap O') \cup [I' \setminus (D \cup O')] \subset (D \cap O) \cup [I \setminus (D \cup O)].$$

Here, $D \cap O = O$, $D \cup O = D$, and $I \setminus D = I$ by Lemma 4.9. Now, suppose to the contrary that (I, O) is not a minimal explanation of A. Then there is an explanation (J, P) of A such that $J \subseteq I$ and $P \subseteq O$ but that $J \neq I$ or $P \neq O$. Similar to the case of (I, O), $D \cap P = P$, $D \cup P = D$, and $J \setminus D = J$. On the other hand, $J \cup P \subset I \cup O$. Using these relations, we have the relation that

$$(D \cap P) \cup [J \setminus (D \cup P)] \subset (D \cap O) \cup [I \setminus (D \cup O)].$$

This contradicts the fact (iii). Therefore, (I, O) is a minimal explanation of A.

The converse direction can also be shown in the same manner. □

The above characterization extends Kakas and Mancarella's [1990] abductive framework of view update. Moreover, our framework provides a mechanism of view update in a flexible manner. This is because we can freely specify dynamic beliefs Γ so that atoms from the extensional base are chosen to prefer some intended update and reduce the non-determinism.

Example 4.11 [Manchandra and Warren, 1988] Let D be a database containing the rules:

$$em(x, z) \leftarrow ed(x, y), dm(y, z),$$
$$ed(sam, publicity),$$
$$dm(sales, john).$$

The first rule says that an employee x has a manager z if x belongs to a department y and the manager of y is z. Suppose that $\mathcal{EB}$ is the set of all extensional atoms having the predicate ed or dm. To insert the view $em(sam, john)$, there are two possible ways to accomplish the update: to insert $dm(publicity, john)$, or to insert $ed(sam, sales)$. If the company allows to make an employee work in two departments but disallows to let a manager supervise two departments, only the second update is permitted. Such a situation is specified in our abductive framework as $\langle D, \{ed(x, y)\} \rangle$. Note that the same effect is achieved in [Manchandra and Warren, 1988] using a framework of dynamic logic programming.

In Definition 4.8, we have assumed that the hypotheses Γ are $\mathcal{EB}$, while, when needed, we can even change some rules in D by putting them into Γ.

4.3 Contradiction Removal

An autoepistemic theory often fails to have a consistent stable expansion. For example, the theory $K_3 = \{ Bp \vee p \}$ has no stable expansion, while $K_4 = \{ \neg p, \ p \}$ has no consistent stable expansion. Here, we address an application of autoepistemic update to resolve contradiction in such theories.

Definition 4.12 Let K be an autoepistemic theory, and $\mathcal{L}_0$ the set of objective formulas in the language. The theory $r(K) = (K \cup I) \setminus O$, where $I, O \subseteq \mathcal{L}_0$, is called a *CR-theory* of K if

1. $r(K)$ is consistent, and
2. for any pair (I', O') such that $(K \cup I') \setminus O'$ is consistent, $I' \subseteq I$ and $O' \subseteq O$ imply that $I' = I$ and $O' = O$.

Notice that $r(K) = K$ if K is consistent.

By definition, the CR-theory $r(K)$ resolves inconsistency in K by minimally introducing or removing appropriate objective formulas in the language. In the following, we formalize the CR-theory in our abductive framework.

Definition 4.13 Let K and Γ be autoepistemic theories. A theory $K' = (K \cup I) \setminus O$, where $I, O \subseteq \Gamma$, is a *most coherent extension* of K (with respect to Γ) if K' is consistent and for any pair (I', O') such that $(K \cup I') \setminus O'$ is consistent, $I' \subseteq I$ and $O' \subseteq O$ imply that $I' = I$ and $O' = O$.

It is easy to see that most coherent extensions in Definition 4.13 generalize CR-theories. That is,

Proposition 4.14 *Let K and $\mathcal{L}_0$ be the same as in Definition 4.12. $r(K)$ is a CR-theory of K iff $r(K)$ is a most coherent extension of K with respect to $\mathcal{L}_0$.*

Note that K is the unique most coherent extension of K iff K is consistent. The next theorem shows that most coherent extensions of K with respect to Γ in general (including CR-theories) can be characterized in the abductive framework $\langle K, \Gamma \rangle$.

Theorem 4.15 *Let $\langle K, \Gamma \rangle$ be any abductive framework. $K' = (K \cup I) \setminus O$, where $I, O \subseteq \Gamma$, is a most coherent extension of K with respect to Γ iff (I, O) is a minimal anti-explanation of $\neg F \wedge F$ with respect to $\langle K, \Gamma \rangle$, where F is any objective formula.*

For the introductory examples, $r(K_3) = \{ Bp \vee p,\ p \}$ with $(I, O) = (\{p\}, \emptyset)$, and $r(K_4) = \{ p \}$ with $(I, O) = (\emptyset, \{\neg p\})$ or $\{ \neg p \}$ with $(I, O) = (\emptyset, \{p\})$, resolve inconsistency of K_3 and K_4, respectively. Note that there are some autoepistemic theories that have no CR-theories but have most coherent extensions. For example, the theory

$$K_5 = \{ p,\ Bp \supset q,\ Bp \supset \neg q,\ \neg Bp \supset r,\ \neg Bp \supset \neg r \}$$

has no CR-theory but has a most coherent extension K_5' with $(I, O) = (\emptyset, \{Bp \supset q\})$. However, the restriction on inserted formulas to $\mathcal{L}_0$ is reasonable. For example, K_3 could become consistent by inserting any of p, $Bq \vee p$, $Br \vee p$, $Bq \vee Br \vee p$, and so on, but in any case the objective formulas of the stable expansion are identical to $cons(\{p\})$.

Morris [1989] provides a method of revising autoepistemic theories having no stable expansion. Given an autoepistemic theory K, he defines a *stable closure* of K as a stable expansion of a minimal augmented theory $G = cons(K \cup I)$, where I is some set of objective formulas such that $K \cup I$ has a stable expansion. Note that every autoepistemic theory has a stable closure. In contrast to ours, objective formulas can only be added to resolve inconsistency in his setting. In fact, Morris's proposal is motivated by dependency-directed backtracking in Doyle's truth maintenance system, which resolves inconsistency by making some disbelieved propositions true. Due to this restriction, an autoepistemic theory may have no consistent stable closure. For example, consider the theory $K_6 = \{ \neg p,\ Bp \vee p \}$ having no stable expansion. The only stable closure of K_6 is $E(\mathcal{L}_0)$. On the other hand, our contradiction removal method provides a CR-theory $r(K_6) = \{ Bp \vee p,\ \ p \}$ with $(I, O) = (\{p\}, \{\neg p\})$ where $r(K_6)$ has the consistent stable expansion $E(\{p\})$.

5 Related Work

It is recognized that nonmonotonic theory update is an important future topic in AI and nonmonotonic reasoning. However, not much work exist on this topic.

There are some work which relate update semantics to abduction. Boutilier [1994] relates abduction to Katsuno and Mendelzon's [199lb] propositions! update semantics, but does not consider nonmonotonic theories ae background theories. Kakas and Mancarella [1990] characterize update semantics through abduction, while their concern is limited to view update in databases.

Marcus and Subrahmanian [1994] recently established the relationship between Fagin *et* al.'s [1983] update and default/autoepistemic logic, but they do not discuss the issue of updating nonmonotonic theories. There are some proposals for removing inconsistency from logic programs with negation as failure. Those approaches in [Pereira *et* a/., 1991; Giordano and Martelli, 1990; Witteveen *et al,* 1994] recover consistency by adding some new formulas, while [Inoue, 1994] discards some beliefs to this effect. In contrast to them, our framework performs update by both inserting and deleting hypotheses based on the extended abductive mechanism.

In the field of theory revision, the *AGM-postulates* [Alchourfon *et al.,* 1985] and their applications to various revision/update systems are thoroughly studied by Katsuno and Mendelzon [I991a; 1991b]. However, those postulates are defined for monotonic propositional theories, and not applicable to our nonmonotonic autoepistemic theory in their present forms. Moreover, many of the revision systems are model-based and deal with *belief sets,* which are closed under logical consequences. By contrast, our approach is formula-based and deals with *belief bases,* which are not necessarily closed under logical consequences, and is syntax-dependent in its nature. In logic programming and deductive databases, formulas included in a theory have their own intended meaning and syntax plays an important role to represent commonsense knowledge. Nebel [I99I] proposes a syntax-based revision system and relates it to some default reasoning systems, but he considers only propositional theories and its applications to logic programming and deductive databases are not addressed.

6 Conclusion

We have proposed a new framework for nonmonotonic theory change. This framework is based on a new form of nonmonotonic abduction, which can explain observations not only by adding some hypotheses to the theory but by retracting some previous hypotheses. With this abductive framework, autoepistemic update was defined for nonmonotonic theory revision and contraction, and then applied to account for view update of deductive databases, first-order theory revision, and contradiction removal for autoepistemic theories. Future work includes devising postulates for nonmonotonic theory change like [Alchourfon *et* a/., 1985; Katsuno and Mendelzon, 1991a; 1991b], developing an efficient mechanism for computing negative and anti- explanations, and investigating connections to update specification languages like [Marek and Truszczynski, 1994].

Our abductive framework is fairly general and can deal with nonmonotonic theories as background theories. The notions of explanations are extended to allow *positive* and *negative explanans* and *anti-explanans.* An inserted formula that changes the world is an *explanandum* sentence, and a contracted formula is an *anti-explanandum* sentence. This extended framework is, we believe, much closer to Peirce's theory of abduction, in which a series of explanatory hypotheses accounting for observations must be revised by experimental testing. The theory of abduction thus relies on the continuous cycle of experiments, observations, hypothesis generation, hypothesis verification, and hypothesis revision.

References

[Abiteboul, 1988] Serge Abiteboul. Updates: a new frontier. In: Proceedings of the 2nd International Conference on Database Theory, Lecture Notes in Computer Science, 326, pages 1-18, Springer, 1988.

[Alchourron et al, 1985] Carlos E. Alchourfon, Peter Gardenfors, and David Makinson. On the logic of theory change: partial meet contraction and revision functions. Journal of Symbolic Logic, 50(2):510-530, 1985.

[Boutilier, 1994] Craig Boutilier. Abduction to plausible causes: an event-based model of belief update. Technical Report 94-9, Department of Computer Science, The University of British Columbia, Vancouver, B. C, March 1994.

[Fagin et al, 1983] Ronald Fagin, Jeffery D. Ullman, and Moshe Y. Vardi. On the semantics of updates in databases (preliminary report). In: Proceedings of the 2nd ACM SIGACT-SIGMOD Symposium on Principles of Database Systems, pages 352-365, 1983.

[Gelfond and Lifschitz, 1991] Michael Gelfond and Vladimir Lifschitz. Classical negation in logic programs and disjunctive databases. New Generation Computing, 9(3,4):365-385, 1991.

[Giordano and Martelli, 1990] Laura Giordano and Alberto Martelli. Generalized Btable models, truth maintenance and conflict resolution. In: Proceedings of the 7th International Conference on Logic Programming, pages 427-441, MIT Press, 1990.

[Inoue, 1994] Katsumi Inoue. Hypothetical reasoning in logic programs. Journal of Logic Programming, 18(3):191-227, 1994.

[Inoue and Sakama, 1994] Katsumi Inoue and Chiaki Sakama. On positive occurrences of negation as failure. In: J. Doyle, E. Sandewall, and P. Torasso (eds.), Principles of Knowledge Representation and Reasoning: Proceedings of the 4th International Conference, pages 293-304, Morgan Kaufmann, 1994.

[Kakas and Mancarella, 1990] A. C. Kakas and P. Mancarella. Database updates through abduction. In: Proceedings of the 16th International Conference on Very Large Databases, pages 650-661, Morgan Kaufmann, 1990.

[Katsuno and Mendelzon, 1991a] Hirofumi Katsuno and Alberto O. Mendelzon. Propositional knowledge base revision and minimal change. Artificial Intelligence, 52:263-294, 1991.

[Katsuno and Mendelzon, 1991b] Hirofumi Katsuno and Alberto O. Mendelzon. On the difference between updating a knowledge base and revising it. In: J. A. Allen, R. Fikes, and E. Sandewall (eds.), Principles of Knowledge Representation and Reasoning: Proceedings of the 2nd International Conference, pages 387-394, Morgan Kaufmann, 1991.

[Konolige, 1992] Kurt Konolige. Ideal introspective belief. In: Proceedings of AAAI-92, pages 635-641, MIT Press, 1992.

[Lifschitz and Schwarz, 1993] Vladimir Lifschitz and Grigori Schwarz. Extended logic programs as autoepistemic theories. In: L. M. Pereira and A. Nerode (eds.), Logic Programming and Non-monotonic Reasoning: Proceedings of the 2nd International Workshop, pages 101-114, MIT Press, 1993.

[Makinson and Gardenfors, 1989] David Makinson and Peter Gardenfors. Relations between the logic of theory change and nonmonotonic reasoning. In: Proceedings of the Workshop on the Logic of Theory Change, Lecture Notes in Artificial Intelligence, 465, pages 185-205, Springer, 1989.

[Manchandra and Warren, 1988] Sanjay Manchandra and David Scott Warren. A logic-based language for database updates. In: J. Minker (ed.), Foundations of Deductive Databases and Logic Programming, pages 363-394, Morgan Kaufmann, 1988.

[Marcus and Subrahmanian, 1994] Sherry Marcus and V. S. Subrahmanian. Relating database updates to nonmonotonic reasoning. Draft manuscript, 1994.

[Marek and Truszczynski, 1994] Victor W. Marek and Miroslaw Truszczynski. Revision specifications by means of programs. In: Proceedings of the 4th European Conference on Logics in AI, Lecture Notes in Artificial Intelligence, 838, pages 122-136, Springer, 1994.

[Moore, 1985] Robert C. Moore. Semantical considerations on nonmonotonic logic. Artificial Intelligence, 25:75-94, 1985.

[Morris, 1989] Paul H. Morris. Autoepistemic stable closures and contradiction resolution. In: Proceedings of the 2nd International Workshop on Non-Monotonic Reasoning, Lecture Notes in Artificial Intelligence, 346, pages 60-73, Springer, 1989.

[Nebel, 1991] Bernhard Nebel. Belief revision and default reasoning: syntax-based approaches. In: J. A. Allen, R. Fikes, and E. Sandewall (eds.), Principles of Knowledge Representation and Reasoning: Proceedings of the 2nd International Conference, pages 417-428, Morgan Kaufmann, 1991.

[Pereira et al., 1991] Luis Moniz Pereira, Joaquim Nunes Aparicio, and Jose Julio Alferes. Contradiction removal within well-founded semantics. In: A. Nerode, W. Marek and V. S. Subrahmanian (eds.), Proceedings of the 1st International Workshop on Logic Programming and Non-monotonic Reasoning, pages 105-119, MIT Press, 1991.

[del Val and Shoham, 1994] Alvaro del Val and Yoav Shoham. A unified view of belief revision and update. Journal of Logic and Computation, 4(5):797-810, 1994.

[Winslett, 1988] Marianne Winslett. Reasoning about action using a possible model approach. In: Proceedings of AAAI-88, pages 89-93, Morgan Kaufmann, 1988.

[Winslett, 1990] Marianne Winslett. Updating Logical Databases. Cambridge Univ. Press, 1990.

[Witteveen et al, 1994] Cees Witteveen, Wiebe van der Hoek, and Hans de Nivelle. Revision of nonmonotonic theories. In: Proceedings of the 4th European Conference on Logics in AI, Lecture Notes in Artificial Intelligence, 838, pages 137-151, Springer, 1994.