On the proper definition of minimality in specialization and theory revision

Stefan Wrobel

GMD (German National Research Center for Computer Science)
FIT.KI, Pf. 1316
W-5205 St. Augustin 1
wrobel@gmdzi.gmd.de

Abstract. A central operation in an incremental learning system is the specialization of an incorrect theory in order exclude incorrect inferences. In this paper, we discuss what properties are to be required from such theory revision operations. In particular, we examine what it should mean for a revision to be minimal. As a surprising result, the seemingly most natural criterion, requiring revisions to produce maximally general correct specializations, leads to a number of serious problems. We therefore propose an alternative interpretation of minimality based on the notion of base revision from theory contraction work, and formally define it as a set of base revision postulates. We then present a revision operator (MBR) that meets these postulates, and shown that it produces the maximally general correct revision satisfying the postulates, i.e., the revisions produced by MBR are indeed minimal in our sense. The operator is implemented and used in KRT, the knowledge revision tool of the MOBAL system.

1 Introduction

An incremental learning system receives a sequence of positive and negative examples from its environment, and always maintains one current hypothesis that is updated after each new input. Two possible cases can arise: If the new input is a positive example (true statement) not covered by the current hypothesis, the hypothesis needs to be *generalized* so that this input can then be explained by the hypothesis. If the new input is a negative example (incorrect statement) erroneously covered by the system's current hypothesis, the system must *specialize* its hypothesis so that the incorrect inference is not produced any more. Ideally, the system should change its current hypothesis *minimally* in order to keep as much as possible of the knowledge it already had.

In this paper, we are concerned with the specialization problem for incremental learning systems that maintain a *first-order theory* as their current hypothesis, i.e., the system learns not only a single clause for one concept, but a set of clauses for a set of interrelated concepts. For single clauses or rules, there is a good understanding of what it means to modify them minimally in order to exclude a negative example — the system should simply perform a minimal specialization (in the logical sense) on the clause. The situation is more difficult

when multiple-clause, multiple concept theories are used as the system's hypothesis, since here, we must not only modify individual clauses, but first decide which clauses to modify. Clearly, in this setting the system should also revise its current hypothesis minimally, but what is the exact definition of minimality in this case? To answer this question Machine Learning can benefit from existing work on *theory revision*, specifically from work on the logic of theory change [Gär88].

In this paper, we for the first time explicitly examine the proper definition of minimality to be used when revising a first-order theory in an incremental learning system. In particular, we will take issue with the hypothesis that minimal revision means minimal specialization of a theory, as has been argued in [MB92]. By relating specialization to work on theory contraction, we will show that such revisions correspond to contraction operations on closed theories, which have very undesirable properties from a computational point of view. We instead propose a more adequate notion of minimality by adapting the well-known set of Gärdenfors postulates for closed theory revision [Gär88] to revisions of theories in an incremental learning system.

We then present the MBR (for "minimal base revision") operator for revision of clausal first-order theories that is shown to meet the revised set of postulates, and is fully implemented in KRT, the knowledge revision tool of the MOBAL system [MWKEss]. It has been successfully used for revision of a set of telecommunication access control rules when faced with a number of incorrect access assignments in an application developed in cooperation with Alcatel Alsthom Recherche, Paris [SMAU93]. In contrast to algorithms previously proposed by [MB92] and [Lin91], our method correctly computes the set of all possible minimal changes to the original knowledge base, and can ensure that lost inferences are recovered when removed statements are re-added to the theory. By using an exception set notation, our method also shows that the use of non-monotonically interpreted predicates, as advocated in [MB92, Lin91], is an unnecessary complexity for minimal theory specialization.

The paper is organized as follows. In the next section (2), we will introduce basic definitions and notations to be used in the rest of the paper. In section 3, we will define the minimal specialization (mgcs) hypothesis, and point out its negative properties by relating to work on closed theory contraction. In section 4, we will present our alternative interpretation of minimality in revision by defining a set of postulates that any revision operation must satisfy if it is to avoid the problems of minimal specialization. In section 5, we then present an operator that implements our notion of minimality, and show that this operator meets our postulates, and actually produces the most general correct revisions that meet the postulates, i.e., is indeed minimal in this sense. Section 6 discusses some related work, and section 7 contains summary and conclusions.

2 Basic definitions

In the following, we will denote with Γ the current knowledge base (inductive hypothesis) of the learning system, where Γ is meant to be a first-order theory in the standard logical syntax as defined eg. in [Men87][1]. We will further assume that standard first-order derivability is used as a consequence operation, denoted by $\vdash$[2], and let $Cn(\Gamma)$ denote the closure of Γ under $\vdash$. If $\Gamma_2 = Cn(\Gamma_1)$, Γ_1 is called a *base* of Γ_2; if $Cn(\Gamma) = \Gamma$, Γ is called *closed*. As a simple example, consider the knowledge base

$$\Gamma_1 = \{p(a), \forall x : p(x) \rightarrow q(x)\}.$$

Assuming that the underlying alphabet contains the predicate symbols p and q, no function symbols, and the constants a and b, we find that

$$\Gamma_2 := Cn(\Gamma_1) = \{p(a), \forall x : p(x) \rightarrow q(x), q(a), p(a) \vee p(b), p(a) \vee \neg p(b), \ldots\}.$$

Γ_1 is a base of Γ_2, and Γ_2 is a closed theory. Finally, by a substitution we mean a possibly empty set of pairs $\sigma = \{v_1/t_1, \ldots, v_n/t_n\}$, where all the v_i are variable symbols and pairwise different, and the t_i are terms. The application of a substitution to a statement s, denoted $s\sigma$, is the statement obtained by replacing in s each occurrence of v_i with t_i for all i.

To more formally define the problem of specializing a theory against a negative example, note that in the general case, this negative example can be any statement about the world that the system currently believes, and is now told to be wrong. Therefore, if we define a *fact* to be a (not necessarily ground)[3] positive or negated atom, we can formally define the specialization or revision problem as follows.

Definition 1 (Revision) Given *a theory Γ and a set of facts $F = \{f_1, \ldots, f_n\}$, find a revised theory $\Gamma \hat{-} F$ ("Γ minus F") such that*

$$Cn(\Gamma \hat{-} F) \cap F = \emptyset.$$

If $F = \{f\}$ is a singleton set we will write $\Gamma \hat{-} f$ instead of $\Gamma \hat{-} \{f\}$.

For simplicity of notation, in the following we will give all results for the singleton set case; as pointed at the end of section 5, they all carry over to the case of multiple simultaneous revisions by performing the obvious modifications on definitions and theorems.

Below, we will also need the generalization/specialization relationship on theories, which as usual we define with respect to derivability.

Definition 2 (Generality) *A theory Γ_1 is said to be* more general *than a theory Γ_2, written $\Gamma_1 \geq_g \Gamma_2$ iff*

$$\Gamma_1 \vdash \Gamma_2.$$

By the definition of Cn, this is equivalent to $\Gamma_1 \geq_g \Gamma_2 \Leftrightarrow Cn(\Gamma_1) \supseteq Cn(\Gamma_2)$.

---

[1] We use $\rightarrow$ instead of $\Rightarrow$ to denote implication, however.

[2] We will usually write $\Gamma \vdash S$ instead of "$\Gamma \vdash s$ for all $s \in S$.

[3] If nonground, variables are assumed to be universally quantified.

3 Minimal specialization and theory contraction

Since the definition of the revision task seriously underconstrains the operator $\dot{-}$, the question we must consider is which further properties to require from the revision result $\Gamma \dot{-} f$. Above, we already pointed out that in general, we want our revisions to be minimal in some sense. In this section, we will examine a seemingly natural instantiation of the notion of minimality that has been proposed in a recent paper by Muggleton and Bain [MB92], who define minimal revisions as *minimal specializations* of a theory. In particular, they define the notion of a maximally general correct specialization as follows.

Definition 3 (mcgs) *Let f be an incorrect statement derived by a theory Γ. A maximally general correct specialization (mgcs) of Γ with respect to f is any theory Γ' such that*

1. $\Gamma \geq_g \Gamma'$
2. $\Gamma' \nvdash f$
3. For all Γ'': If $\Gamma \geq_g \Gamma'' >_g \Gamma'$, then $\Gamma'' \vdash f$.

In other words, a *mgcs* is any theory that is a specialization of the original theory, and does not have any correct supersets, i.e., supersets that do not imply the fact to be removed. We will often refer to the *mgcs* simply as the *minimal specialization* of Γ with respect to f.

To evaluate the properties of this notion of minimal revision, we can draw upon some well-known results from work on closed theory contraction [Gär88, Neb89]. As we will see, maximally general correct specializations are a special case of the operations examined there. In the next section, we will briefly summarize some important results, following the exposition in [Neb89, ch. 6].

3.1 Closed theory contraction

The goal of work on closed theory contraction was a knowledge-level examination of theory change operations. Consequently, in this work, it is generally assumed that the theories under consideration are *closed*. Theory contraction then refers to the same task that we have called theory revision above. One of the best-known results from theory contraction work is the set of *Gärdenfors postulates*, which are a set of minimal constraints claimed to be necessary for every sensible revision operation. The six major postulates are summarized in table 1, for a closed theory Γ and a fact f to be removed. These postulates are mostly unspectacular and describe reasonable properties one would expect from a revised theory. Postulate 1 simply requires the revised theory to be closed, which if necessary can be ensured by applying Cn. Postulate 2 is more interesting, as it requires the revised theory to be a subset of the original theory, thus ruling out revision operations that generalize some part of the theory. Postulates 3 and 4 simply capture the definition of the revision task, requiring the operation to leave the theory alone if f was not in the theory, and requiring it to effectively

---

1. $\Gamma \dot{-} f$ is a closed theory *(closure)*.
2. $\Gamma \dot{-} f \subseteq \Gamma$ *(inclusion)*.
3. If $f \notin \Gamma$, then $\Gamma \dot{-} f = \Gamma$ *(vacuity)*.
4. If $f \notin Cn(\emptyset)$, then $f \notin \Gamma \dot{-} f$ *(success)*.
5. If $Cn(f) = Cn(g)$, then $\Gamma \dot{-} f = \Gamma \dot{-} g$ *(preservation)*.
6. $\Gamma \subseteq Cn(\Gamma \dot{-} f \cup \{f\})$ *(recovery)*.

Table 1. The main Gärdenfors postulates for closed theory contraction

remove f if f is not a tautology (which is always the case if f is a fact). Postulate 5 requires the preservation of semantic equivalence.

The final postulate, recovery, is the most interesting one from the standpoint of machine learning, as it addresses a likely scenario for an incremental learning system. The recovery postulate requires that if we modify a theory to exclude an inference f, and later find out that f was true after all, we should be able to obtain at least the same set of inferences again that we used to have before removing f in the first place. It thus acts as a lower bound on the revisions we may perform, and excludes for instance the trivial empty theory as a revision result.

Interestingly, as pointed out in [Neb89], all possible revision operations that meet the above postulates can be defined with respect to the set of maximal correct subsets of a closed theory.

Definition 4 (Maximal correct subsets) *If Γ is a theory, and f a statement to be removed, the family of maximal subsets of Γ not implying f, denoted $\Gamma \downarrow f$ (pronounced "Γ down f" or "Γ less f") is defined by:*

$$\Gamma \downarrow f := \{\Gamma' \subseteq \Gamma | f \notin Cn(\Gamma') \text{ and for all } \Gamma'' : \text{ if } \Gamma' \subset \Gamma'' \subseteq \Gamma \text{ then } f \in Cn(\Gamma'')\}.$$

As an example, consider the closed theory Γ_2 from the example in section 2 above, i.e.,

$$\Gamma_2 = \{p(a), \forall x : p(x) \rightarrow q(x), q(a), p(a) \vee p(b), p(a) \vee \neg p(b), \ldots\}.$$

If we want to remove $q(a)$ from this theory, some of the maximally correct subsets are:

$$\Gamma_2 \downarrow q(a) = \left\{ \begin{array}{l} \{p(a), p(a) \vee p(b), p(a) \vee \neg p(b), \ldots\} \\ \{\forall x : p(x) \rightarrow q(x), p(a) \vee p(b), \ldots\} \\ \{\forall x : p(x) \rightarrow q(x), p(a) \vee \neg p(b), \ldots\} \\ \ldots \end{array} \right\}$$

The following theorem (first shown in [AGM85]) allows us to express any revision operation on closed theories that meets the Gärdenfors postulates in one common form.

Theorem 1 $\dot{-}$ *is a revision operation that meets the Gärdenfors postulates 1–6 if and only if there is a selection function* $\gamma : \Gamma \downarrow f \mapsto \gamma(\Gamma \downarrow f) \subseteq \Gamma \downarrow f$ *such that*

$$\Gamma \dot{-} f = \begin{cases} \bigcap \gamma(\Gamma \downarrow f) & \text{if } f \notin Cn(\emptyset) \\ \Gamma & \text{otherwise.} \end{cases}$$

In other words, any revision operation meeting postulates 1 through 6 can be defined by specifying a function γ that returns a subset of $\Gamma \downarrow f$, the set of all maximally correct subsets of Γ, and computing the intersection of these sets. Depending on the size of the set returned by γ, the resulting operations are given different names:

Definition 5 (Partial meet, full meet, maxi-choice contractions) *All revision operations of the form defined in theorem 1 are called* partial meet contractions. *If γ returns a singleton, the operation is called a* maxi-choice contraction, *if γ simply returns $\Gamma \downarrow f$, the operation is called* full-meet contraction.

3.2 Mgcs as maxi-choice contractions

Returning to minimal specializations resp. maximally general correct specializations (mgcs), we can see that they correspond precisely to the maxi-choice contractions on a closed theory as defined above.

Theorem 2 *Given a (not necessarily closed) theory Γ, and statement to be removed f, the* mgcs *of Γ with respect to f are exactly the results of all possible maxi-choice contractions on $Cn(\Gamma)$, i.e., the members of the set $Cn(\Gamma) \downarrow f$.*

Proof. The proof follows trivially from definition 4, according to which $Cn(\Gamma) \downarrow f =$

$$\{\Gamma' \subseteq Cn(\Gamma) | f \notin Cn(\Gamma') \text{ and for all } \Gamma'' : \text{if } \Gamma' \subset \Gamma'' \subseteq Cn(\Gamma) \text{ then } x \in Cn(\Gamma'')\},$$

which is a reformulation of the definition of *mgcs*. □

This correspondence, unfortunately, means that minimal specializations inherit all the known undesirable properties of maxi-choice contractions as revision operations, as pointed out by [Neb89] and others:

Nonfinite representation. Since minimal specialization revisions are equivalent to maxi-choice operations on closed theories, they also produce closed theories. In general, it is impossible to determine whether a closed theory has a finite axiomatization, which is a necessary prerequisite for the use of minimal specialization for practical theory revision.

Loss of reason maintenance. An operation on the closure of a theory does not take into account which statements were derived from which others. For instance, if from Γ_1 (section 2), we were to remove $p(a)$ by minimal specialization, one of the possible minimal specializations would be

$$mgcs(\Gamma_1, p(a)) = Cn(\Gamma_1) \downarrow p(a)$$
$$= \{\{\forall x : p(x) \to q(x), q(a), p(a) \vee p(b), \ldots\}, \ldots\},$$

i.e., we would remove the antecedent $p(a)$, but keep statements that were derived from it such as $q(a)$.

Besserwissers. Last, for a maxi-choice contraction $\dot{-}$, we know (theorem 6.2 from [Neb89] foll. [AM82]) that for any proposition g:

either $g \in Cn(\Gamma \dot{-} f \cup \neg f)$, or $\neg g \in Cn(\Gamma \dot{-} f \cup \neg f)$.

This means that if a learning system finds out that instead of f, as currently believed, $\neg f$ is true, and uses minimal specialization to remove f before adding $\neg f$, the system will have miraculously completed its theory, which is certainly not desirable[4].

4 Minimal base revisions

We can thus conclude that an interpretation of minimal revisions as minimal specializations, or *mgcs* in the terminology of [MB92], is not an advantageous strategy, since it is an operation that can be performed only on closed theories, resulting in the drawbacks enumerated above. Based on the base contraction postulates in [Neb89], we have therefore developed a revised set of revision postulates that applies to revision operations on non-closed first-order theories; these postulates are shown in table 2.

---

1. $\Gamma \dot{-} f \subseteq \Gamma \cup \{g' | \exists g \in \Gamma \text{ such that } g \geq_g g'\}$ (*minimal syntactic distance*).
2. $Cn(\Gamma \dot{-} f) \subseteq Cn(\Gamma)$ (*inclusion*).
3. If $f \notin Cn(\Gamma)$, then $\Gamma \dot{-} f = \Gamma$ (*vacuity*).
4. If $f \notin Cn(\emptyset)$, then $f \notin Cn(\Gamma \dot{-} f)$ (*success*).
5. If $Cn(f) = Cn(g)$, then $Cn(\Gamma \dot{-} f) = Cn(\Gamma \dot{-} g)$ (*preservation*).
6. $Cn(\Gamma) \subseteq Cn(\Gamma \dot{-} f \cup \{f\})$ (*recovery*).

Table 2. Revised set of postulates for revision of theory bases

The most important difference between the closed theory postulates and the ones defined here is the first postulate, which together with the recovery postulate, expresses a new notion of minimal revision for theory bases. According to

---
[4] The term besserwisser is due to [Gär88].

the first postulate, a revision is minimal if the new theory contains only statements that had been in the old theory, and perhaps other statements that are specializations of statements in the original theory. Note that minimal specialization revisions in general do not meet this new first postulate, as their results will contain elements from the closure of Γ that were not originally in Γ nor are specializations of single statements in Γ.

If we assume that our theories are in clausal form, as we will from now on, we can define $\geq_g$ for individual clauses simply as θ-subsumption [Plo70]:

$$f \geq_g f' \text{ iff there is a substitution } \theta \text{ such that } f\theta \subseteq f'.$$

This definition ensures that indeed the revised knowledge base is syntactically close to the original knowledge base, since new clauses must be derived by substitution or added literals from existing clauses.

5   A minimal base revision operator

To show that the above set of revised postulates indeed allows useful theory revision operations, we will now present a revision operator for clausal theories that meets all of these postulates. This operator, which we will call MBR, for *minimal base revision*, and denote by $\stackrel{\sim}{-}$, has been implemented and is used in KRT, the knowledge revision tool of the MOBAL system [MWKEss], to perform knowledge revision. In order to define $\stackrel{\sim}{-}$, we need to define the *derivation tree* of a fact in a theory. As usual, we will assume that $\vdash$ is implemented as refutation proofs by resolution.

Definition 6 (Derivation) *Let Γ be a theory, and $f \in Cn(\Gamma)$ a factual query finitely refutable from Γ by resolution. The derivation of f in Γ, $\Delta(f, \Gamma)$, is the pair*

$$\Delta(f, \Gamma) := (f, S),$$

where S, the supports *of f, is the following set of triples:*

$$S := \{(C, \sigma, A) | \exists C \in \Gamma \wedge \exists g_1, \ldots, g_n \in Cn(\Gamma) : \quad C \text{ resolves with}$$
$$\{g_1, \ldots, g_n\} \text{ using substitution } \sigma \text{ to produce } f\}.$$

A, the antecedents *of S, are recursively defined as follows:*

$$A := \{\Delta(g_1, \Gamma), \ldots, \Delta(g_n, \Gamma)\}.$$

Wherever Γ is clear from context, we will simply write $\Delta(f)$ instead of $\Delta(f, \Gamma)$.

Since different refutation proofs of f may share the same clauses and substitutions, the (C, σ, A) triples representing clause applications may repeat in a derivation structure, meaning that a derivation is best understood as a directed acyclic graph. In the following, we will assume that for any inferences f to be removed, $\Delta(f, \Gamma)$ is finite, which may not be the case in theories with function symbols. From a derivation, a standard *proof tree* can be obtained by beginning

at the root, choosing one of its supports, and then repeating this recursively for the derivation subtrees rooted at the antecedents of this supports. By varying the choices in all possible ways, the set of all proof trees can be obtained. As an example of a derivation, consider the theory

$$\Gamma_1 = \{p(a), \forall x : p(x) \rightarrow q(x)\}.$$

defined in section 2. Dropping the quantification symbol, and using capitalized variables instead, this theory is written as

$$\Gamma_1 = \{p(a), p(X) \rightarrow q(X)\}.$$

Let us assume that q(a) is to be removed from this theory. There is only one way of proving q(a), so

$$\Delta(q(a), \Gamma_1) = (q(a),$$
$$\{(p(X) \rightarrow q(X), \{X/a\},$$
$$\{(p(a),$$
$$\{(p(a), \emptyset, \emptyset)\}) \}) \}),$$

Using this definition of a derivation, we can define the clause application set of a theory Γ with respect to a fact f.

Definition 7 (Application set) *Let Γ be a theory and $f \in Cn(\Gamma)$ a fact. We define the* clause application set *of Γ with respect to f as:*

$$\Pi(f, \Gamma) := \{(C, \sigma) | \exists S = (C, \sigma, A) \text{ somewhere in } \Delta(f, \Gamma)\}.$$

Instead of $(C, \emptyset)$ (where $\emptyset$ in this context is the empty substitution), we often simply write C in $\Pi(f, \Gamma)$.

In our example, the clause application set is

$$\Pi(q(a), \Gamma_1) = \{(p(a), \emptyset), (p(X) \rightarrow q(X), \{X/a\})\},$$

and we define for later use abbreviations for the two nonempty subsets of this application set:

$$P_1 := \{(p(a), \emptyset)\}$$
$$P_2 := \{(p(X) \rightarrow q(X), \{X/a\})\}$$

These application sets are needed so that each clause application in the proof of an offending fact can be individually kept or removed; otherwise, we would unnecessarily overspecialize in cases where one rule is used several times. Consequently, below we will define a version of $\Gamma \downarrow f$ that can work on application sets instead of on theories. First, however, we need to define substitution and instance sets.

Definition 8 (Substitution and Instance sets) *Let C be a clause in a theory Γ, with variables $\mathrm{vars}(C) = \{X_1, \ldots, X_n\}$, let $f \in Cn(\Gamma)$ be a fact, and $P \subseteq \Pi(f, \Gamma)$. The substitution set of C with respect to P is defined as follows. We say that one substitution is more general than another, $\sigma_1 \geq_g \sigma_2$, iff there is a substitution θ such that $\sigma_1 \theta = \sigma_2$. Let*

$$\Sigma_0(C, P) := \{\sigma | (C, \sigma) \in P\}$$

denote all substitutions used with C in P. Then

$$\Sigma(C, P) \subseteq \Sigma_0(C, P)$$

is any set such that:

- *for any $\sigma \in \Sigma_0(C, P)$, there is a $\sigma' \in \Sigma(C, P)$ such that $\sigma' \geq_g \sigma$.*
- *for no $\sigma \in \Sigma(C, P)$, there is a $\sigma' \in \Sigma(C, P)$ such that $\sigma' \geq_g \sigma$.*

For non-ground clauses ($n \geq 1$), the instance set of C with respect to P is defined as

$$I(C, P) := \{(X_1, \ldots, X_n)\sigma | \sigma \in \Sigma(C, P)\}.$$

We should note that the construction of $\Sigma(C, P)$ ensures that there are no redundant substitutions in this set. Continuing with our example, we see that

$$\Sigma(p(a), P_1) = \{\emptyset\}$$
$$\Sigma(p(X) \to q(X), P_2) = \{\{X/a\}\}$$
$$\Sigma(p(a), P_2) = \Sigma(p(X) \to q(X), P_1) = \emptyset$$
$$I(p(X) \to q(X), P_1) = \emptyset$$
$$I(p(X) \to q(X), P_2) = \{(a)\}.$$

We can now define which theory corresponds to an application set:

Definition 9 (Corresponding theory) *Let Γ be a theory and $f \in Cn(\Gamma)$ a fact, and $P \subseteq \Pi(f, \Gamma)$. If we define the clauses occurring in P as*

$$C(P) := \{C | (C, \sigma) \in P\},$$

we can define the theory corresponding to P as:

$$\begin{aligned}
\Gamma_\Pi(P) := \; & (\Gamma \backslash C(P)) \\
& \cup \{\mathrm{vars}(C) \notin I(C, P) \diamond C \mid C \in C(P) \text{ and } C \text{ nonground}\} \\
& \cup \{C\sigma | (C, \sigma) \in P\}.
\end{aligned}$$

Here, $L_0 \diamond C$ is a shorthand notation for the addition of a premise, i.e., if C is a clause of the form $L_1 \& \cdots \& L_n \to L_{n+1}$, then

$$L_0 \diamond C := L_0 \& L_1 \& \cdots \& L_n \to L_{n+1}.$$

$\mathrm{vars}(C)$ denotes the tuple of variables of the clause.

We can now define the promised version of $\Gamma \downarrow f$ that can work on application sets instead of on theories.

Definition 10 (Maximal application, Minimal removal sets) *Let Γ be a theory, and f a fact to be removed. The* set of maximal application sets *of Γ with respect to f is defined as:*

$$\Gamma\downarrow_\Pi f := \Pi(f,\Gamma)\downarrow f := \{P \subseteq \Pi(f,\Gamma)|f \notin Cn(\Gamma_\Pi(P)) \text{ and for all } P' :$$
$$\text{if } P \subset P' \subseteq \Pi(f,\Gamma) \text{ then } f \in Cn(\Gamma_\Pi(P'))\}.$$

The complement of $\Gamma\downarrow_\Pi f$ is called the set of *minimal removal sets, and defined by*

$$\overline{\Gamma\downarrow_\Pi f} := \overline{\Pi(f,\Gamma)\downarrow f} := \{\Pi(f,\Gamma)\backslash P|P \in \Gamma\downarrow_\Pi f\}.$$

In our example, there are two maximal correct application subsets, which are each others complement, so

$$\Gamma_1\downarrow_\Pi q(a) = \overline{\Gamma_1\downarrow_\Pi q(a)} = \{P_1, P_2\}.$$

As our last preliminary, we now define sets of clauses that are added to a theory to replace any clause, or more precisely, any application of a clause, that needs to be removed.

Definition 11 (Add set) *For a clause C with variables* $\text{vars}(C) = \{X_1,\ldots,X_n\}$, $n \geq 1$, *in a theory Γ, a fact f to be removed, and $P \subseteq \Pi(f,\Gamma)$ such that $C \in C(P)$, i.e. occurs in P, define*

$$add(C,f,P) := \{(X_1,\ldots,X_n) \notin I(C,P) \diamond C\} \cup \{(f \diamond C)\sigma|\sigma \in \Sigma(C,P)\}^5.$$

If $n = 0$, i.e., C is ground, $\Sigma(C,P) = \{\emptyset\}$, so we define

$$add(C,f,P) := \{f \diamond C\}.$$

This definition assumes that set membership ($\in$) is a predefined predicate that is properly handled by the proof procedure with sets of both ground and non-ground tuples, i.e. as if $\in$ were defined by

$(T_1,\ldots,T_n) \in S$ iff there exists $T \in S$ and a substitution σ so that $T\sigma = (T_1,\ldots,T_n)$.

For instance, $(a) \in \{(b)\}$ will evaluate to false, whereas $(f(a)) \in \{(f(X))\}$ will evaluate to true[6]. In our example, the add sets for our two statements in Γ_1 would be:

$$add(p(a),q(a),P_1) = \{q(a)\rightarrow p(a)\},$$
$$add(p(X)\rightarrow q(X),q(a),P_2)$$
$$= \{(X) \notin \{(a)\}\& p(X)\rightarrow q(X), q(a)\& p(a)\rightarrow q(a)\}$$
$$= \{(X) \notin \{(a)\}\& p(X)\rightarrow q(X)\}.$$

We can now define the minimal base revision (MBR) operator $\stackrel{\frown}{-}$.

---

[5] Instead of the set of $(f \diamond C)\sigma$, it would be possible to simply use $(f\sigma \diamond C)$, but this would produce redundancy in the revised theory.

[6] For unary tuples, we usually omit the parentheses. In KRT/MOBAL, the exception sets are represented in an equivalent, but syntactically slightly different form called a *support set* [Wro88]. In this paper, we will stick with the form as just defined for simplicity.

Definition 12 (MBR operator $\hat{-}$) *Let Γ be a knowledge base, f be a fact to be removed, and γ a selection function on $\Gamma \downarrow_\Pi f$. If $\hat{P} := \gamma(\Gamma \downarrow_\Pi f)$ denotes the chosen maximally correct application subset of $\Pi(f, \Gamma)$, and $\overline{P} := \Pi(f, \Gamma) \backslash \hat{P}$ its complement, we can define the* minimal base revision (MBR) *operator $\hat{-}$ as follows:*

$$\Gamma \hat{-} f := \begin{cases} \Gamma \backslash C(\overline{P}) \;\cup\; \bigcup_{C \in C(\overline{P})} add(C, f, \overline{P}) \text{ if } f \in Cn(\Gamma) \\ \Gamma \hspace{6cm} \text{otherwise.} \end{cases}$$

The above definition assumes that we are given some choice function γ that returns a single element from $\Gamma \downarrow_\Pi f$. For the purposes of this paper, we may simply assume that γ is implemented by user queries, and that $\Gamma \downarrow_\Pi f$ is computed by a top-down search through the subset lattice of Γ. We should point out, however, that $\Gamma \downarrow_\Pi f$ (or rather, its complement) can be computed efficiently with a recursive procedure based on the derivation tree of the fact to be removed. In [Wro93], we describe this procedure in detail, and also develop a two-tiered confidence model on the basis of which γ can be defined. The implementation of $\hat{-}$ in KRT/MOBAL in addition uses a confidence-ranked beam search to ensure bounded computation times for $\Gamma \downarrow_\Pi f$.

In the above example, the MBR operation would produce the following results. If the first element of $\Gamma_1 \downarrow_\Pi q(a)$, P_1, were chosen by γ, we would have to remove P_2, so the resulting new theory would be

$$\begin{aligned} \Gamma_1 \hat{-} q(a) &= \{p(a), (X) \notin \{(a)\} \& \; p(X) \to q(X), q(a) \& \; p(a) \to q(a)\}; \\ &= \{p(a), (X) \notin \{(a)\} \& \; p(X) \to q(X)\}; \end{aligned}$$

if the second element, P_2 were chosen, we would remove P_1, resulting in

$$\Gamma_1 \hat{-} q(a) = \{q(a) \to p(a), p(X) \to q(X)\}.$$

As another example involving a non-ground fact to be removed, consider the theory (shown in rule form with capitalized variables for readability)

$$\Gamma_3 = \begin{cases} f1 : p(X) \\ r1 : p(X) \to q(X) \\ r2 : q(X) \to r(X) \end{cases}$$

from which we want to remove q(f(Y)). In this example, we see that the minimal removal sets are

$$\overline{\Gamma_3 \downarrow_\Pi q(f(Y))} = \{\{(r1, \{X/f(Y)\})\}, \{(f1, \{X/f(Y)\})\}\}.$$

The two theories that result from these choices are:

$$\Gamma_3 \hat{-} q(f(Y)) = \begin{cases} f1 : \; p(X) \\ r11 : \; X \notin \{f(Y)\} \& \; p(X) \to q(X) \\ [r12 : \; q(f(Y)) \& \; p(f(Y)) \to q(f(Y))] \\ r2 : \; q(X) \to r(X) \end{cases}$$

where $r12$ is a tautology and would be omitted, and

$$\Gamma_3 \hat{-} q(f(Y)) = \left\{ \begin{array}{l} f11 : X \notin \{f(Y)\} \rightarrow p(X) \\ f12 : q(f(Y)) \rightarrow p(f(Y)) \\ r1 : \quad p(X) \rightarrow q(X) \\ r2 : \quad q(X) \rightarrow r(X) \end{array} \right\}$$

respectively. To see why the revision operator must consider individual clause applications in deciding about minimal removal sets, consider the example of

$$\Gamma_4 = \left\{ \begin{array}{l} p(X) \\ p(a) \& p(b) \rightarrow r(a) \end{array} \right\}$$

from which we want to remove $r(a)$. If we choose to modify $p(X)$ (or rather, one or more of its applications), $\hat{-}$ will correctly produce either one of

$$\left\{ \begin{array}{l} (X) \notin \{(a)\} \rightarrow p(X) \\ r(a) \rightarrow p(a) \\ p(a) \& p(b) \rightarrow r(a) \end{array} \right\} \quad \text{or} \quad \left\{ \begin{array}{l} (X) \notin \{(b)\} \rightarrow p(X) \\ r(a) \rightarrow p(b) \\ p(a) \& p(b) \rightarrow r(a) \end{array} \right\}$$

whereas a revision operation not based on individual clause applications would have to produce the overly specific

$$\left\{ \begin{array}{l} (X) \notin \{(a), (b)\} \rightarrow p(X) \\ r(a) \rightarrow p(a) \\ r(a) \rightarrow p(b) \\ p(a) \& p(b) \rightarrow r(a) \end{array} \right\}$$

We now prove that indeed $\hat{-}$ meets the set of postulates for base revision operations defined in table 2.

Theorem 3 $\hat{-}$ *is a theory revision operation that meets the base revision postulates (1) through (6) from table 2.*

Proof. We show each postulate individually. For the exceptional case where $f \notin Cn(\Gamma)$, the postulates are trivially true, so in the following we will assume $f \in Cn(\Gamma)$.

(1). Since $\Gamma \backslash C(\overline{P}) \subseteq \Gamma$ anyway, we only need to show that for any removed clause $C \in C(\overline{P})$, $add(C, f, \overline{P})$ only contains statements that are θ-subsumed by elements of Γ. Looking at the definition of add, we see that new statements are defined by adding literals to and/or instantiating existing statements, so this is indeed true.

(2). True as an immediate consequence of (1).

(3). True by definition of $\hat{-}$.

(4). By the definition of $\Gamma \downarrow_\Pi f$, f is not derivable from the theories correspon-
ding to its members. It thus remains to verify that this is not changed by the
additional statements from *add*. This is easy to see, because the additional pre-
mises added to removed statements ensure that all substitutions which could be
used to prove f are excluded; the other additional statements can be used only
if f is present.

(5). Trivially true since for factual queries, $Cn(f) = Cn(g)$ implies $f = g$.

(6). We can easily show that $\Gamma \subseteq Cn(\Gamma \dot{-} f \cup \{f\})$. For non-removed statements,
this is trivially true. For each nonground removed statement C, we find that

$$Cn(add(C, f, \overline{P}) \cup \{f\})$$
$$= Cn(\{(X_1, \ldots, X_n) \notin I(C, \overline{P}) \diamond C\} \cup \{(f \diamond C)\sigma | \sigma \in \Sigma(C, \overline{P})\} \cup \{f\})$$
$$= Cn(\{(X_1, \ldots, X_n) \notin I(C, \overline{P}) \diamond C\} \cup \{C\sigma | \sigma \in \Sigma(C, \overline{P})\})$$
$$= Cn(C),$$

and similarly for ground statements. □

The reader may be curious how $\dot{-}$ relates to the closed theory contraction
operations defined above, and in particular, whether using a singleton choice
function γ will not produce the same besserwisser effect that we criticized for
minimal specialization revisions. Indeed this is not the case, since on the level
of closed theories, this singleton choice actually corresponds to a set of subsets
of the closed theory, as shown by the following theorem, the proof of which can
be found in [Wro93].

Theorem 4 *For a theory Γ and a fact f, $\dot{-}$ is equivalent to a partial meet*
contraction on the theory $\Gamma' := \Gamma_\Pi(\Pi(f, \Gamma))$, i.e., the theory that results when all
clauses are split into their individual applications (cf. definition 9). In particular,
given a selection function γ,

$$Cn(\Gamma' \dot{-} f) = \begin{cases} \bigcap \gamma_{Cn}(Cn(\Gamma') \downarrow f) & \text{if } f \notin Cn(\emptyset) \\ Cn(\Gamma) & \text{otherwise} \end{cases},$$

where

$$\gamma_{Cn}(Cn(\Gamma') \downarrow f) := \{B \in (Cn(\Gamma') \downarrow f) \mid \exists C \in \Gamma_\Pi(\gamma(\Gamma \downarrow_\Pi f)) \text{such that } C \subseteq B\}.$$

Knowing that $\dot{-}$ meets the set of revision postulates from table 2, the re-
maining open question is whether it truly captures the essence of minimality as
expressed by these postulates, i.e., whether it produces the maximally general
base revisions consistent with these postulates. Indeed this is the case.

Theorem 5 (Minimality) *Let $\dot{-}'$ any base revision operation that also meets*
the base revision postulates. Then for any theory Γ, and any fact f,

$$\Gamma \dot{-} f \geq_g \Gamma \dot{-}' f.$$

Proof. By postulate (1), $\Gamma \overset{\sim}{-}' f$ must consist of a subset of Γ, plus perhaps some added statements. With a proper selection function, we can make $\overset{\sim}{-}$ choose the same subset. As for the added statements, we know they must be specializations of existing statements. We thus must only show that $add(C, f, \overline{P})$, where $\overline{P} = \overline{\gamma(\Gamma \downarrow_\Pi f)}$, is a minimal specialization of C to complete the proof. So assume that there is a statement C' such that $C \geq_g C' >_g add(C, f, \overline{P})$. That is, there must be some substitution (inference) possible with both C and C' that is not admitted by $add(C, f, \overline{P})$. Since $add(C, f, \overline{P})$ excludes precisely the substitutions mentioned in the minimal removal set $\overline{P}$, however, we know that none of them may be readmitted without rederiving f again. Thus C' is not a correct specialization of C. □

We conclude by pointing out that $\overset{\sim}{-}$ can very simply be extended to multiple parallel revisions. So let $F = \{f_1, \dots, f_n\}$ a set of facts to be removed. If we define

$$\Delta(F, \Gamma) := (\Delta(f_n, \Gamma), \dots, \Delta(f_n, \Gamma)),$$

and replace f by F in all other definitions, we see that all of the methods still apply, and that all proofs still carry through as the reader will be able to verify.

6   Related work

The approach presented here is an elaboration of the knowledge revision method used in the knowledge acquisition and machine learning system BLIP, a predecessor of MOBAL. As described in [Wro88, Wro89], BLIP already used exception sets to produce minimal specializations of individual rules. As in MOBAL, these exception sets were represented as support sets, a simple form of which was first proposed in [EHR83]. In other respects, BLIP was seriously lacking in contrast to the method described here that is used by KRT/MOBAL. In particular, the computation of removal sets was incomplete in many cases, and no formal characterization was available. Furthermore, BLIP could not work on multiple revisions at the same time, and could not ensure recovery (postulate 6).

In section 3, we already discussed the problems of the minimal specialization hypothesis that was proposed in [MB92]. In the same paper, the authors also present a specialization algorithm based on the introduction of non-monotonically interpreted premises with new predicates. Thus, for example, if the substitutions $\{\{X/a\}, \{X/b\}\}$ were to be excluded from $p(X) \to q(X)$, the algorithm of Muggleton and Bain would produce the theory

$$p(X) \wedge not(c1(X)) \to q(X); c1(a); c1(b),$$

where $c1$ is a new predicate, and *not* is interpreted as negation by failure. As can easily be seen, this is a notational variant of the exception set method used to specialize clauses used in BLIP and KRT/MOBAL. Thus, the introduction of non-monotonically interpreted predicates just adds unnecessary complexity, as it is not necessary for correct minimal specialization of clauses. The algorithm

of [MB92] always selects to modify those clauses that have directly resolved with the fact that is to be removed. Nonetheless, contrary to what is implied in [MB92], this is insufficient to guarantee minimal specialization, as can be seen by considering our example theory Γ_3. In this example, when modifying the clause $r1$ that resolved with $q(f(Y))$, we necessarily lose the inference $r(f(Y))$ that was possible with the original theory. Nonetheless, this fact cannot be used to derive $q(f(Y))$, so it would be in a minimal specialization.

Our method of using exception lists to specialize individual clauses has also been adopted in the algorithm proposed in [Lin91]. This paper also introduces the important notion of learning in a growing language: specialization operations should be such that when a new constant is added to the language anything provable about the new constant with the original theory should also be provable from the specialization. As Ling points out, the exception methods ensures that this is the case. He also points to an important problem that was also recognized in [Wro88, Wro89]: if exclusion is used as the only specialization operator on individual clauses, we may build up long (possibly infinite) exception lists. In KRT, this problem is addressed through a user specified *plausibility criterion* that defines when further specialization is necessary to get rid of an overly long exception list. KRT then applies a number of specialization operators that specialize further, essentially by introducing new premise literals on existing variables of the clause (cf. [Wro88, Wro89]). Ling instead uses a *complete* set of refinement operators, i.e., capable of producing all specializations of a clause, and simply replaces a clause by all of its specializations. This guarantees minimal specialization and thus identification in the limit, but brings with it the undesirable properties of minimal specializations as spelled out in section 3.

MIS [Sha83] was one of the first first-order learning systems to include theory revision. In MIS, however, minimality of revision was not a concern, since subsequent generalization steps were relied upon to fix up a theory that had become overspecialized. Nonetheless, the MIS approach is highly relevant to the work presented here, since its backtracing algorithm offers a way of determining with a minimum number of user queries which possible revision to choose, i.e., it offers one particular way of implementing γ, the choice function among minimal revisions. As pointed out above (but not described in this paper), KRT uses a different method of implementing γ that relies on a two-tiered model of confidence in statements, and proposes to the user the revision that would entail a minimal loss of confidence. Evidently, this could easily combined with a backtracing strategy. The backtracing strategy of MIS is also used in the interactive learning programs MARVIN [SB86] and CLINT [DeR91] to recover from overgeneralizations that lead to incorrectly covered negative examples.

7 Conclusion

In this paper, we have discussed the question of what properties are to be required from the results of theory revision operations, i.e., operations that specialize the current hypothesis of a learning system in order to remove an erroneously

covered inference. In particular, we have examined how the notion of minimality (or "Occam's razor") should be interpreted in this context. As a surprising result, it has turned out that a most natural restriction from the standpoint of Machine Learning, namely the use of minimal specializations, has very undesirable properties: because it leads to closed theories, loses any notion of reason maintenance, and when removing a statement to add its negation, results in complete theories. We have therefore developed an alternative notion of minimality, and formalized this notion in the form of a set of base revision postulates derived from similar postulates by Gärdenfors [Gär88] and Nebel [Neb89]. As the perhaps central contribution of the paper, we have described a minimal base revision operator MBR ("$\dot{-}$") that produces revisions meeting all of the revision postulates; moreover, it produces the maximally general correct revisions meeting the postulates, and is thus really minimal in our alternative sense.

In precisely specifying how to perform minimal base contractions, the work described here can be used as a basis for approaching a number of interesting questions of a larger scope in theory revision in incremental learning systems. The first one concerns the role of the recovery postulate, which we have adopted from Gärdenfors' theory change work. While it is a desirable property for a revision algorithm to be able to give a recovery guarantee to a user, it is nonetheless not clear in what circumstances a learning system (or a user) would want to actually make use of it, since it adds clauses to the theory that are largely uninteresting as long as the removed fact is not added back in. A second question concerns the question of when to perform minimal specialization in the sense as defined here, and when to perform non-minimal specialization, eg. in order to obtain a more succinct theory. The plausibility criterion currently used in KRT is only a rudimentary answer to this problem. Last, and certainly not least, is the open question of whether there is something to be gained by combining specialization and generalizing revisions into one general and theory revision operation, as argued by [Emd89]. Such an operation would most likely be driven mainly not by criteria such as completeness or correctness, but by "scientific" notions about the structural qualities of a good theory.

Acknowledgements

References

[AGM85]   C. E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: Partial meet contraction and revision functions. *The Journal of*

Symbolic Logic, 50:510 – 530, 1985.

[AM82]   C. E. Alchourrón and D. Makinson. On the logic of theory change: contraction functions and their associated revision functions. *Theoria*, 48:14 – 37, 1982.

[DeR91]   Luc DeRaedt. *Interactive Concept-Learning*. PhD thesis, Kath. Univ. Leuven, Leuven, Belgium, February 1991.

[EHR83]   Werner Emde, Christopher U. Habel, and Claus-Rainer Rollinger. The discovery of the equator or concept driven learning. In *IJCAI-83*, pages 455 – 458, San Mateo, CA, 1983. Morgan Kaufman.

[Emd89]   Werner Emde. An inference engine for representing multiple theories. In Katharina Morik, editor, *Knowledge Representation and Organization in Machine Learning*, pages 148–176. Springer Verlag, Berlin, New York, 1989.

[Gär88]   Peter Gärdenfors. *Knowledge in Flux — Modeling the Dynamics of Epistemic States*. MIT Press, Cambridge, MA, 1988.

[Lin91]   Charles Ling. Non-monotonic specialization. In *Proc. Inductive Logic Programming Workshop, Portugal*, 1991.

[MB92]   Stephen Muggleton and Michael Bain. Non-monotonic learning. In Stephen Muggleton, editor, *Inductive Logic Programming*. Academic Press, London, New York, 1992.

[Men87]   Elliot Mendelson. *Introduction to Mathematical Logic*. Wadsworth & Brooks, Belmont, CA, third edition, 1987.

[MWKEss]   K. Morik, S. Wrobel, J.U. Kietz, and Werner Emde. *Knowledge Acquisition and Machine Learning: Theory, Methods, and Applications*. Academic Press, London, New York, in press. To appear.

[Neb89]   Bernhard Nebel. *Reasoning and Revision in Hybrid Representation Systems*. Springer Verlag, Berlin, New York, 1989. Doctoral Dissertation.

[Plo70]   Gordon D. Plotkin. A note on inductive generalization. In B. Meltzer and D. Michie, editors, *Machine Intelligence 5*, chapter 8, pages 153 – 163. Edinburgh Univ. Press, Edinburgh, 1970.

[SB86]   Claude Sammut and Ranan B. Banerji. Learning concepts by asking questions. In R.S. Michalski, J.G. Carbonell, and T.M. Mitchell, editors, *Machine Learning — An Artificial Intelligence Approach*, volume II, chapter 7, pages 167 – 191. Morgan Kaufman, San Mateo, CA, 1986.

[Sha83]   Ehud Y. Shapiro. *Algorithmic Program Debugging*. ACM Distinguished Doctoral Dissertations. The MIT Press, Cambridge, Mass., 1983.

[SMAU93]   E. Sommer, K. Morik, J-M. André, and M. Uszynski. What online machine learning can do for knowledge acquisition — a case study. To appear, 1993.

[Wro88]   Stefan Wrobel. Automatic representation adjustment in an observational discovery system. In D. Sleeman, editor, *Proc. of the 3rd Europ. Working Session on Learning*, pages 253 – 262, London, 1988. Pitman.

[Wro89]   Stefan Wrobel. Demand-driven concept formation. In Katharina Morik, editor, *Knowledge Representation and Organization in Machine Learning*, pages 289–319. Springer Verlag, Berlin, New York, 1989.

[Wro93]   Stefan Wrobel. Representation adjustment by demand-driven concept formation. in preparation, 1993.