

A Many-Sorted Resolution based on an Extension of a First-Order Language*

K. B. Irani and D. G. Shin

Robot Systems Division
Center for Research on Integrated Manufacturing
College of Engineering, The University of Michigan,
Ann Arbor, Mi. 48109

ABSTRACT

A type of problem is first identified which may occur when a resolution scheme is applied to a many-sorted theory. To avoid such a problem, an extension of the first-order language called *one-sorted language with aggregate variables* is introduced. Aggregate variables allow the introduction of range-restricted variables dynamically in the structure which is expanded by definitions. This allows the introduction of a new resolution scheme called *UWR-resolution*. It can be shown that the UWR-resolution is more efficient than the existing schemes.

1. Introduction

Within the field of theorem proving, the advantages of a many-sorted logic are well known [Henschen, 1972 and Cohn, 1983]. More recently a sound theoretical foundation for many-sorted resolution has been established in [Walther, 1983, 1984a] and the power of a many-sorted resolution has been demonstrated by an example in [Walther, 1984b]. However, because of its associated fixed structure, sometimes not as much advantage can be derived from the use of a many-sorted calculus as one would like to get. This is illustrated by the following example.

Example 1.1 Let x_a, x_b, x_c, x_d , and x_e be the variables ranging over the sorts A, B, C, D , and E , respectively, where $D \subset B, D \subset C, E \subset A, E \subset B$, and $E \subset C$. Suppose the theory to be refuted is given by:

- (1) $\forall x_a \forall x_b (P(x_b) \cup Q(x_b, f^d(x_b)) \cup R(x_a, x_b)),$
- (2) $\forall x_c \neg P(x_c),$
- (3) $\forall x_e \forall x_c \neg Q(x_e, x_c),$
- (4) $\forall x_c \neg R(g^e(x_c), x_c),$

where $f^d(x_b)$ and $g^e(x_c)$ are Skolem functions whose ranges are restricted to D and E , respectively. Suppose (1) and (2) are resolved, because x_b of $P(x_b)$ in (1) and x_c of $\neg P(x_c)$ in (2) are unifiable over D . The two clauses can be resolved using the most general unifier (mgu) $\theta = \{y_d/x_b, y_d/x_c\}$ where y_d ranges over D . The deduction is then

- (5) $Q(y_d, f^d(y_d)) \cup R(x_a, y_d) \quad (1)+(2),$
- (6) $Q(y_d, f^d(y_d)) \quad (4)+(5).$

It is noticed that (6) cannot be resolved with any other clauses, not even with (3) because no sort is known to be a subsort of D and E . A dead end has been reached. (5) is a useless resolvent, so is (6).

Had there been another sort $G = B \cap C$, x_b of $P(x_b)$ in (1) and x_c of $\neg P(x_c)$ in (2) could have been unified over the sort G giving the deduction

- (5') $Q(y_g, f^d(y_g)) \cup R(x_a, y_g) \quad (1)+(2),$
- (6') $Q(y_g, f^d(y_g)) \quad (4)+(5'),$

where y_g ranges over G . The clause (6') can be further resolved with (3) resulting in the empty clause $\square$. There is no dead end here. It is observed that the need for G is not known until (1) and (2) are attempted to be resolved. In an ordinary many-sorted language, one cannot add a new sort whenever a need arises without altering the sort structure.

To alleviate such a situation, an extension of a one-sorted language called *one-sorted language with aggregate variables* (L_{Σ}^1) is proposed into which a many-sorted theory can be translated and which may be dynamically extended to bypass the problem illustrated above. Previously in [Shin and Irani, 1984], the concept of embedding aggregate variables (a.v.) into a many-sorted language (L_m) was reported, and its resulting language was called *many-sorted language with a.v.*. Here a.v.'s are embedded in a one-sorted language.

a.v.'s allow the introduction of range-restricted variables dynamically without revising the a priori fixed structure. Using the dynamic range-restricting nature of an a.v., an efficient many-sorted resolution scheme called *UWR-resolution* is presented which is designed to avoid generating useless resolvents as illustrated in the above example. The completeness of UWR-resolution can be shown. The efficiency of UWR-resolution is discussed.

2. One-sorted Language with a.v. L_{Σ}^1

2.1. Syntax of L_{Σ}^1

In L_{Σ}^1 two types of variables, called simple variables and a.v.'s, are available. A simple variable of L_{Σ}^1 is the same as the ordinary variable of one-sorted language. Unlike a simple variable, an a.v. is syntactically an ordinary sort variable, but semantically a variable whose range of interpretation is restricted by a unary relation rather than to a sort universe. Formally stated, an a.v. is of the form x^{EP_i} in which x^{EP_i} ranges over the unary relation indicated by the unary predicate symbol P_i .

Let J, L , and K be respectively a relation index set, a function index set and a constant index set. In addition, let I be an index set for some unary relations. Let λ and ξ be functions such that $\lambda: J \rightarrow N^+$ and $\xi: L \rightarrow N^+$, where N^+ is the set of positive integers.

Definition 2.1 A *one-sorted language with a.v.* L_{Σ}^1 then consists of the following: (1) parentheses $(,)$ and a symbol Σ ;

*This work was supported in part by Air Force Office Scientific Research under contract F49620-82-C-0089.

(2) constant symbol C_k , for each $k \in K$; (3) simple variables $x_1, \dots, x_m, \dots$, and a.v.'s $x_1^{\Sigma P_1}, \dots, x_n^{\Sigma P_n}$, for each $i \in I$, where P_i is a unary predicate symbol; (4) a $\lambda(j)$ -ary predicate symbol R_j , for each $j \in J$; (5) a $\xi(l)$ -ary function symbol F_l , for each $l \in L$; (6) logical connectives $\neg$, and $\rightarrow$; and (7) a universal quantifier $\forall$.

Based on this language, the *terms* of L_Σ^1 are defined as usual except that each variable is now either a simple variable or an a.v.. The set, $Form(L_\Sigma^1)$, of *well-formed formulas* of L_Σ^1 , is defined as usual. The definable syntactic objects $\cup$, $\cap$, $\models$ and $\exists$, and the standard notions such as *sentences* are also introduced in the usual way.

2.2. Interpretation of L_Σ^1

A structure is needed to interpret each formula in L_Σ^1 . Let OS_a be a structure for L_Σ^1 . Then $OS_a = \langle \Omega, \{P_i\}_{i \in I}, \{R_j\}_{j \in J}, \{F_l\}_{l \in L}, \{C_k\}_{k \in K} \rangle^*$ where Ω is the universe of OS_a ; P_i is a unary relation $P_i \subseteq \Omega$; R_j is a $\lambda(j)$ -ary relation $R_j \subseteq \Omega^{\lambda(j)}$; F_l is a $\xi(l)$ -ary function $F_l: \Omega^{\xi(l)} \rightarrow \Omega$; and a distinguished element C_k is an element of Ω , for each $k \in K$. The interpretation of a formula in the structure OS_a then requires a variable assignment function s as follows:

Definition 2.2 For a set V of variables of L_Σ^1 and the universe Ω of structure OS_a , s is an assignment function $s: V \rightarrow \Omega$ such that for a simple variable x , $s(x) = a$, where $a \in \Omega$; and for an aggregate variable $x^{\Sigma P_i}$, $s(x^{\Sigma P_i}) = a$, where $a \in P_i$.

Assignment function for the terms of L_Σ^1 is defined as usual. For notational convenience symbol s is also used for the assignment for the terms. The validity of each formula is determined by the following interpretation rules.

Definition 2.3 For $R_j(t_0, \dots, t_{\lambda(j)})$, $\psi_1, \psi_2 \in Form(L_\Sigma^1)$, where t_i 's are terms, the satisfaction of the formulas with respect to s in OS_a is defined by,

- (1) $\models_{OS_a} R_j(t_0, \dots, t_{\lambda(j)})[s]$ iff $\langle s(t_0), \dots, s(t_{\lambda(j)}) \rangle \in R_j$,
- (2) $\models_{OS_a} \neg \psi_1[s]$ iff $\not\models_{OS_a} \psi_1[s]$,
- (3) $\models_{OS_a} \psi_1 \rightarrow \psi_2[s]$ iff if $\models_{OS_a} \psi_1[s]$ then $\models_{OS_a} \psi_2[s]$,
- (4) For a simple variable x , $\models_{OS_a} \forall x \psi_1[s]$ iff for any $a \in \Omega$, $\models_{OS_a} \psi_1[s(x|a)]$,
- (5) For an aggregate variable $x^{\Sigma P_i}$, $\models_{OS_a} \forall x^{\Sigma P_i} \psi_1[s]$ iff for any $a \in P_i$, $\models_{OS_a} \psi_1[s(x^{\Sigma P_i}|a)]$,

where for variables v_m and v_k ,

$$s(v_m | a)(v_k) = \begin{cases} s(v_k) & \text{if } v_m \neq v_k \\ a & \text{if } v_m = v_k \end{cases}$$

As a corollary to the definition, the interpretations of $\cup$, $\cap$, $\models$ and $\exists$ can also be easily defined.

2.3. Σ -Extensibility of L_Σ^1

L_Σ^1 is as convenient as L_m in abbreviating the relatively expressions in a one-sorted language into more compact

* From next section on, " $\cdot$ " on a symbol is omitted as long as the meaning of the symbol is unambiguous.

forms. If a formula σ_m in L_m is of the form

$$\sigma_m = \forall x_i \text{ --- } x_i \text{ ---},$$

where x_i is a sort variable belonging to some sort S_i , then σ_m can be syntactically translated into σ_Σ^* in L_Σ^1

$$\sigma_\Sigma^* = \forall x^{\Sigma P_i} \text{ --- } x^{\Sigma P_i} \text{ ---},$$

where the unary relation intended by P_i is identical to the sort S_i . During the translation, L_Σ^1 is augmented with P_i and accordingly the structure for L_Σ^1 , say $OS_a(L_\Sigma^1)^*$, can be constructed from the many-sorted structure for L_m , say $MS(L_m)$. The following theorem can be shown (proofs for the theorems given in this paper can be found in [Shin, 1985]):

Theorem 2.1 A sentence σ_m in L_m is true in $MS(L_m)$ iff σ_Σ^* in L_Σ^1 is true in $OS_a(L_\Sigma^1)^*$.

The power of L_Σ^1 over L_m lies in the fact that whereas in the latter a sort variable with new sort may not be introduced, in the former a variable whose range is restricted to a subset of the universe Ω can be introduced as needed in its extension. Let T_Σ be a theory in L_Σ^1 and let a formula $\phi_\Sigma \in T_\Sigma$ be of the form

$$\phi_\Sigma = \forall x (S_1(x) \cap S_2(x) \rightarrow \psi(x)).$$

In L_Σ^1 , in order to introduce a variable ranging over $S_1 \cap S_2$, all that must be done is to add a new unary predicate symbol S_k to L_Σ^1 , abbreviate ϕ_Σ by $\forall x^{\Sigma S_k} \psi(x^{\Sigma S_k})$, and augment T_Σ by the defining axiom $\forall x (S_k(x) \models S_1(x) \cap S_2(x))$. The extended language $L_\Sigma^{1'}$ is called a Σ -extension of L_Σ^1 and the augmented theory T_Σ' , a Σ -extension of T_Σ . For the semantics of the new predicate symbols in $L_\Sigma^{1'}$, it can be shown that there is a unique expansion by definition of OS_a , say OS_a' , which is a model of T_Σ' . OS_a' is called an *expansion by Σ -definition* of OS_a . This characteristic of L_Σ^1 is called Σ -extensibility. The following theorem establishes the validity of Σ -extensibility of L_Σ^1 :

Theorem 2.2 For any $\psi' \in T_\Sigma'$, there is a $\psi \in T_\Sigma$ such that for any assignment function s ,

$$\models_{OS_a'} \psi' [s] \iff \models_{OS_a} \psi [s].$$

3. UWR-resolution

The problem identified in Section 1, namely, the generation of useless resolvents that lead to dead ends, occurs only when a many-sorted theory of a certain class is refuted by a resolution scheme. For instance, for the many-sorted theories with tree structure stated in [Walther, 1984a], this problem never occurs. When the tree constraint is lifted, however, this problem may appear. The conditions under which the problem may arise are explained below, this time in terms of L_Σ^1 .

Let $Ran(t)$ stand for the range associated with a term t . Given a set of unary relations P , let a set of immediate predecessor of a relation $P_i \in P$, denoted by $IM(P_i)$, be defined by $IM(P_i) = \{P_j | P_j \in P, P_j \subset P_i, \text{ and } P_j \subseteq P_i \subseteq P_i \text{ then either } P_j = P_i \text{ or } P_i = P_j\}$. A situation may occur in which two variables v_i and v_j have $|IM(Ran(v_i)) \cap IM(Ran(v_j))| > 1$. Hence there are possibly more than one range over which v_i can be unified with v_j . In fact, if x_k is a variable such that $Ran(x_k) = P_k$ and $P_k \in IM(Ran(v_i)) \cap IM(Ran(v_j))$, then any substitution $\theta = \{x_k/v_i, x_k/v_j\}$ is a legitimate mgu of $\{v_i, v_j\}$, since $\theta v_i = \theta v_j$. There are, therefore, as many mgus for

$\{v_i, v_j\}$ as $|IM(Ran(v_i)) \cap IM(Ran(v_j))|$. Not all the resolvents generated using each of these mgus are useful.

A way to remedy the situation is proposed in the following. First, a few new notions are introduced. A *wr-substitution component* is any expression of the form t/v , where v is a variable and t is a term different from v satisfying $Ran(t) \subseteq Ran(v)$. A *wr-subpair* is a set of wr-substitution components $\{t/v_i, t/v_j\}$ satisfying, (1) $Ran(v_i) \not\subseteq Ran(v_j)$ and $Ran(v_j) \not\subseteq Ran(v_i)$, (2) $|IM(Ran(v_i)) \cap IM(Ran(v_j))| > 1$, and (3) $Ran(t) = Ran(v_i) \cap Ran(v_j)$. A *wr-substitution* is a set of wr-substitution components which possibly contains one or more wr-subpairs. A *wr-resolvent* is a resolvent which is generated by using a wr-substitution as a unifier.

The Σ -extensibility of L_Σ^1 plays the central role in introducing wr-subpairs. From the definition of a wr-subpair, it is clear that wr-resolvents can only be expressed in an extended language of L_Σ^1 . To be more specific, let a theory in L_Σ^1 be formalized by an ordered pair $\langle OA, T_\Sigma \rangle$ where OA is a set of ordering axioms expressed with the symbol " $\subset$ " such that $P_i \subset P_j$ means $\forall x (P_i(x) \rightarrow P_j(x))$ and T_Σ is a set of the nonlogical axioms of the theory expressed in L_Σ^1 . The following is an example:

Example 3.1. Consider the following $\langle OA, T_\Sigma \rangle$:

- $$\begin{aligned} OA : & (1) D \subset B, D \subset C \\ & (2) E \subset B, E \subset C \\ T_\Sigma : & (3) \forall x^{\Sigma B} (P(x^{\Sigma B}) \cup Q(x^{\Sigma B}, f^d(x^{\Sigma B}))) \\ & (4) \forall x^{\Sigma C} \neg P(x^{\Sigma C}) \\ & (5) \forall x^{\Sigma E} \forall x^{\Sigma C} \neg Q(x^{\Sigma E}, x^{\Sigma C}) \end{aligned}$$

The following is done: For $x^{\Sigma B}$ of $P(x^{\Sigma B})$ in (3) and $x^{\Sigma C}$ of $\neg P(x^{\Sigma C})$ in (4), a wr-subpair $\{x^{\Sigma K}/x^{\Sigma B}, x^{\Sigma K}/x^{\Sigma C}\}$ is introduced with L_Σ^1 being extended by a unary predicate symbol, say K , where $\forall x (K(x) \subseteq B(x) \cap C(x))$. The extension of L_Σ^1 requires $\langle OA, T_\Sigma \rangle$ to be also extended, i.e., OA is augmented to OA^+ by the ordering axioms such as (2^+) below. (6) is derived as the wr-resolvent of (3) and (4). Finally, (6) is resolved with (5) resulting in $\square$ as follows:

- $$\begin{aligned} (2^+) & K \subset B, K \subset C, D \subset K, E \subset K \\ (6) & Q(x^{\Sigma K}, f^d(x^{\Sigma K})) \quad (3)+(4) \\ (7) & \square \quad (5)+(6) \end{aligned}$$

It is not difficult to see that the range of t in a wr-subpair $\{t/v_i, t/v_j\}$, i.e., $Ran(t) = Ran(v_i) \cap Ran(v_j)$, is the weakest range over which $\{v_i, v_j\}$ can be unified -- weakest in the sense that if $P_w = Ran(t)$, then there is no P_i such that $P_w \subset P_i$ and v_i and v_j are still unifiable over P_i . For this reason, the unification stated above is called **unification over the weakest range** and the resolution involving such unification is called **UWR-resolution**. The idea behind UWR-resolution is therefore to subsume all the possible unifications by one unification over the weakest possible range.

The completeness of UWR-resolution must be proved. Given $\langle OA, T_\Sigma \rangle$, let $R_W(T_\Sigma)$ be the set of all clauses consisting of members of T_Σ and the resolvents (including wr-resolvents) of members of T_Σ (similar to Robinson's resolution operator [Robinson, 1965]). Also let $R_W^n(T_\Sigma)$ be defined so that for each $n \geq 0$, $R_W^0(T_\Sigma) = T_\Sigma$ and $R_W^{n+1}(T_\Sigma) = R_W(R_W^n(T_\Sigma))$. Completeness theorem states:

Theorem 3.1 $\langle OA, T_\Sigma \rangle$ is unsatisfiable if and only if $R_W^n(T_\Sigma)$ contains $\square$, for some $n \geq 0$.

In fact, when the UWR-resolution is applied, $\langle OA, T_\Sigma \rangle$ is refuted not in L_Σ^1 but in its extended language L_Σ^1 . Therefore it further must be justified that when $R_W^n(T_\Sigma)$ contains $\square$, for some $n \geq 0$, whether the unsatisfiability of $\langle OA, T_\Sigma \rangle$ can be truly shown in L_Σ^1 without extending its vocabulary. Let $R_\Sigma(\cdot)$ stand for an identical situation as $R_W(\cdot)$ except that in the former no wr-resolvents are generated. In this way the unsatisfiability of $\langle OA, T_\Sigma \rangle$ can be shown without extending L_Σ^1 . The following theorem can be shown:

Theorem 3.2 Given $\langle OA, T_\Sigma \rangle$, $R_W^n(T_\Sigma)$ contains $\square$, for some $n \geq 0$, if and only if $R_\Sigma^m(T_\Sigma)$ contains $\square$, for some $m \geq 0$.

4. Efficiency of UWR-resolution

One way to discuss the efficiency of UWR-resolution is to compare $R_W(\cdot)$ and $R_\Sigma(\cdot)$ defined previously. The following facts show the efficiency of UWR-resolution:

Lemma 4.1 Given $\langle OA, T_\Sigma \rangle$, if n is the smallest non-negative integer for which $R_W^n(T_\Sigma)$ contains $\square$ and m is the smallest non-negative integer for which $R_\Sigma^m(T_\Sigma)$ contains $\square$, then $n = m$.

Lemma 4.2 Given $\langle OA, T_\Sigma \rangle$, for each $i \geq 0$ $|R_W^{i+1}(T_\Sigma) - R_W^i(T_\Sigma)| \leq |R_\Sigma^{i+1}(T_\Sigma) - R_\Sigma^i(T_\Sigma)|$.

Finally, the following theorem can be proved:

Theorem 4.3 Given $\langle OA, T_\Sigma \rangle$, if n is the smallest non-negative integer for which $R_W^n(T_\Sigma)$ and $R_\Sigma^n(T_\Sigma)$ both contain $\square$, then $|R_W^n(T_\Sigma)| \leq |R_\Sigma^n(T_\Sigma)|$.

5. Conclusion

The use of aggregate variables have been demonstrated in a new approach, namely, UWR-resolution, for proving theorems. It has been shown how such an approach avoids the problems encountered in the use of a many-sorted language.

REFERENCES

- [1] Cohn, A. G. "Improving the Expressiveness of Many-Sorted Logic", *Proc. of the 3rd NCAI*, Washington, 1983.
- [2] Henschen, L. J. "N-Sorted Logic for Automatic Theorem Proving in Higher-Order Logic", *Proc. of ACM Conference*, Boston, 1972.
- [3] Robinson, J. A. "A Machine-Oriented Logic Based on the Resolution Principle", *JACM*, Vol. 12, No. 1, 1965.
- [4] Shin, D. G. "An Extension of a First-Order Language and Its Applications" Ph.D. Dissertation, University of Michigan, Ann Arbor, 1985.
- [5] Shin, D. G. and Irani, K. B. "Knowledge Representation using an Extension of a Many-sorted Language", *Proc. of the 1st CAIA*, Denver, 1984.
- [6] Walther, C. "A Many-sorted Calculus Based on Resolution and Paramodulation", *Proc. of the 8th IJCAI*, Karlsruhe, 1983.
- [7] Walther, C. "Unification in Many-sorted Theories", *Proc. of the 6th ECAI*, Pisa, 1984a.
- [8] Walther, C. "A Mechanical Solution of Schubert's Streamroller by Many-Sorted Resolution", *Proc. of the 4th NCAI*, Austin, 1984b.