Circumscriptive Theories: A Logic-Based Framework for Knowledge Representation (Preliminary Report)

Vladimir Lifschitz

Computer Science Department
Stanford University
Stanford, CA 94305

Abstract

The use of circumscription for formalizing commonsense knowledge and reasoning requires that a *circumscription policy* be selected for each particular application: we should specify which predicates are circumscribed, which predicates and functions are allowed to vary, what priorities between the circumscribed predicates are established, etc. The circumscription policy is usually described either informally or using suitable metamathematical notation. In this paper we propose a simple and general formalism which permits describing circumscription policies by *axioms*, included in the knowledge base along with the axioms describing the objects of reasoning. This method allows us to formalize some important forms of metalevel reasoning in the circumscriptive theory itself.

1. Introduction

The *logic approach* to the problem of knowledge representation, proposed by John McCarthy (1960), stresses the analogy between a knowledge base and an axiomatic theory. Knowledge about the world can be expressed by sentences in a logical language, and an intelligent program should be able to derive new facts from the facts already known, as a mathematician can derive new mathematical results from the facts already proved.

Further research has shown that formal theories like those used for the formalization of mathematics are not adequate for representing some important forms of commonsense knowledge. The facts that serve as a basis for *default reasoning* require more powerful representational languages. Several extensions of the classical concept of a first-order theory have been proposed to resolve this difficulty. We concentrate here on one of these extensions, the concept of *circumscription* (McCarthy 1980, 1986).

Here is a simple blocks world example illustrating McCarthy's approach to formalizing default reasoning. If there is no information to the contrary, we assume that a given block is located on the table and that its color is white. Block B is red. We want a formalization of these assumptions to allow us to conclude by default that all blocks other than B are white (since B is the only block which is known to be an exception) and that all blocks are on the table (since no information to the contrary is available). We can express the given assumptions using two *abnormality predicates* $ab1$ and $ab2$, as follows:

$$\neg ab1\ x \land block\ x \supset ontable\ x, \tag{1}$$

$$\neg ab2\ x \land block\ x \supset white\ x \tag{2}$$

(the blocks that are not abnormal are on the table; the blocks that are not abnormal in a certain other sense are white). The axiom set will also include the formulas

$$block\ B, \quad red\ B, \quad \neg(white\ x \land red\ x). \tag{3}$$

Axioms (1)—(3) are not sufficiently strong for justifying the desired results about the positions and colors of blocks, because they do not say whether there are few or many "abnormal" objects. McCarthy's method consists in *circumscribing* $ab1$ and $ab2$, i.e., assuming their "minimality" subject to the restrictions expressed by the axioms.

There are several different minimality conditions that can be applied in conjunction with a given axiom set. Each kind of minimization corresponds to a different "circumscription policy". In the existing literature on applications of circumscription, these "policies" are described either informally, or using some kind of metamathematical notation, or by establishing a standard convention, as in "simple circumscriptive theories" of (McCarthy 1986).

It has been suggested, on the other hand, that circumscription policies may be described by *axioms* included in the knowledge base along with the axioms describing the objects of reasoning (Lifschitz 1986), (Perlis 1987). We propose in this paper a simple but powerful formalism for expressing such "policy axioms", which leads us to the concept of a *circumscriptive theory*. Our circumscriptive theories are similar to McCarthy's simple circumscriptive theories in the sense that such a theory is completely determined by its axioms, without any additional metamathematical specifications. But this formalism is substantially more expressive; for instance, we

This research was partially supported by DARPA under Contract N0039-82-C-0250.

can axiomatically describe *priorities* in the sense of (McCarthy 1986).

In this preliminary report we discuss only a special case of circumscriptive theories: theories with "propositional" policies. Most applications of circumscription discussed in the literature use policies of this kind. The final version of the paper will present the general case and will also contain proofs of theorems and the discussion of methods used for determining the effect of circumscription in the examples.

The paper does not assume familiarity with previous work on circumscription. A better understanding of the motivation behind the theory of circumscription can be gained by reading (McCarthy 1980, 1986).

2. Theories with Propositional Policies

The language of a circumscriptive theory is defined in the same way as a first-order language, i.e., by a finite set $\mathcal{F}$ of function constants and a finite set $\mathcal{P}$ of predicate constants. (We treat object constants as 0-ary function constants.)

Consider an m-ary predicate constant $P \in \mathcal{P}$. In any model of a given axiom set, P is interpreted as a mapping from U^m, where U is the universe of the model, into $\{false, true\}$. By the minimality of P at a point $(x_1, \ldots, x_m)$ we understand the impossibility of replacing the value of P at that point by a smaller value (i.e., changing it from $true$ to $false$) without losing the properties expressed in the axioms.[1] To make this condition precise, we should specify which additional changes in the values of P and in the values of other predicates or functions, if any, we are allowed to make along with changing $P(x_1, \ldots, x_m)$ to $false$. In other words, for any n-ary function or predicate constant $C \in \mathcal{F} \cup \mathcal{P}$ and any point $(y_1, \ldots, y_n)$, a circumscription policy should define whether the value $C(y_1, \ldots, y_n)$ is allowed to vary in the process of minimizing $P(x_1, \ldots, x_m)$.

This can be done formally by introducing, for each predicate constant $P \in \mathcal{P}$ and each constant $C \in \mathcal{F} \cup \mathcal{P}$, an additional predicate constant V_{PC}, whose arity is the sum of the arities of P and C. "Policy axioms" will contain subformulas of the form $V_{PC}(x_1, \ldots, x_m, y_1, \ldots, y_n)$ expressing that, when P is minimized at point $(x_1, \ldots, x_m)$, C may be varied at $(y_1, \ldots, y_n)$. In simple applications, however, we think of each predicate V_{PC} as being either identically true or identically false, so that the propositional symbol corresponding to the universal closure of this formula will suffice. In this preliminary report we discuss only such cases.[2]

Accordingly, we build *formulas* of a circumscriptive theory from the symbols included in $\mathcal{F}$ and $\mathcal{P}$ and the additional *propositional* constants V_{PC} ($P \in \mathcal{P}$, $C \in \mathcal{F} \cup \mathcal{P}$) using the symbolism of second-order predicate logic. (Thus formulas may include function and predicate variables of any arities; we need such variables for expressing minimality.) V_{PC} reads: *C is varied as P is minimized.*

A *circumscriptive theory* T is defined by a finite set of formulas, *axioms*. We will sometimes identify T with the conjunction of (the universal closures of) its axioms. Thus T can be viewed as a sentence. If $\mathcal{F} \cup \mathcal{P} = \{C_1, \ldots, C_l\}$ then we will also write this sentence as $T(C_1, \ldots, C_l)$.

3. The Semantics of Circumscriptive Theories

To define the semantics of circumscriptive theories, we will construct, for every predicate constant $P \in \mathcal{P}$, a sentence which expresses the minimality of P. Let $c_1, \ldots, c_l$ be distinct variables similar to $C_1, \ldots, C_l$ (i.e., if C_i is an n-ary function (predicate) constant then c_i is an n-ary function (predicate) variable). Recall that P is among the constants $C_1, \ldots, C_l$, so that there is a corresponding predicate variable p in the list $c_1, \ldots, c_l$. Then the *minimality condition* for P in T is

$$\neg\exists \overline{x} c_1 \ldots c_l [P(\overline{x}) \wedge \neg p(\overline{x}) \wedge \bigwedge_{i=1}^{l} (\neg V_{PC_i} \supset c_i = C_i) \wedge T(c_1, \ldots, c_l)], \qquad (4)$$

where $\overline{x}$ is a tuple of distinct object variables. This formula will be denoted by Min_P. It says that it is impossible to change the interpretations of the constants $C_1, \ldots, C_l$ so that the value of P at some point $\overline{x}$ will change from $true$ to $false$, the interpretations of the constants C_i which are not allowed to vary as P is minimized will remain the same, and the new interpretations of $C_1, \ldots, C_l$ will still satisfy T.

For example, if the language of T has two unary predicate constants, P and Q, and no other predicate or function constants then Min_P is

$$\neg\exists x p q [P(x) \wedge \neg p(x) \wedge (\neg V_{PP} \supset p = P) \wedge (\neg V_{PQ} \supset q = Q) \wedge T(p, q)]. \qquad (4')$$

[1] This understanding of minimality is in the spirit of the "pointwise" approach to circumscription (Lifschitz 1986).

[2] The generality afforded by treating V_{PC} as a predicate is needed for describing some circumscriptions with multiple minima, as in Example 2 from (McCarthy 1980), and for formalizing temporal minimization, as in Section 6 of (Lifschitz 1986).

A *model* of a circumscriptive theory T is any model M of the axioms of T which satisfies Min_P for each $P \in \mathcal{P}$. A *theorem* of T is any sentence which is true in every model of T.

We have defined theorems in model-theoretic terms, not in terms of proofs. In view of the incompleteness of second-order logic, a definition based on deduction in second-order predicate calculus would not be equivalent to the one given above. We will develop the theory entirely in model-theoretic terms. In particular, the expressions "A implies B", "B follows from A", where A, B are sentences or sets of sentences, will mean that every model of A is a model of B.

The definition of a model given above includes a minimality condition for *each* predicate constant P in the language. This looks like a serious limitation: in applications, it is often desirable to minimize only some of the available predicates. But minimizing a predicate P remains nominal unless the circumscription policy at least allows us to vary P itself. It is easy to see, for instance, that the assumption $\neg V_{PP}$ makes the first 3 conjunctive terms in (4′) inconsistent, and thus makes the whole formula trivially true. We can use the axiom V_{PP} to say that P is in fact among the predicates which we want to minimize.[3]

The minimality conditions Min_P have a simple model-theoretic meaning. Denote the interpretation of a symbol C in a model M by $M[\![C]\!]$. In particular, for each propositional constant V_{PC}, $M[\![V_{PC}]\!]$ is a truth value, *true* or *false*. We are interested in the models of the axioms of T with a fixed universe U. Let $Mod(T,U)$ be the set of all such models. For every predicate constant $P \in \mathcal{P}$ and every $\xi \in U^m$, where m is the arity of P, we define a reflexive and transitive relation (preorder) $\leq^{P\xi}$ on $Mod(T,U)$ as follows: $M_1 \leq^{P\xi} M_2$ if

(i) $M_1[\![V_{QC}]\!] = M_2[\![V_{QC}]\!]$ for all $Q \in \mathcal{P}$, $C \in \mathcal{F} \cup \mathcal{P}$,

(ii) for all $C \in \mathcal{F} \cup \mathcal{P}$, if $M_1[\![V_{PC}]\!] = false$ then $M_1[\![C]\!] = M_2[\![C]\!]$,

(iii) $M_1[\![P]\!](\xi) \leq M_2[\![P]\!](\xi)$.

Symbol $\leq$ in part (iii) of this definition refers to the usual ordering of truth values ($false < true$). Notice that, in view of (i), $M_1[\![V_{PC}]\!]$ in (ii) can be equivalently replaced by $M_2[\![V_{PC}]\!]$.

Proposition 1. *A model $M \in Mod(T,U)$ is a model of T iff it is minimal in $Mod(T,U)$ relative to each preorder $\leq^{P\xi}$.*

[3] This use of V_{PP} was suggested by John McCarthy.

4. Example

To formalize the blocks world example from the introduction, we take

$$\mathcal{F} = \{B\},\ \mathcal{P} = \{block, ontable, white, red, ab1, ab2\}.$$

Formulas of a circumscriptive theory with these function and predicate constants may also contain $6 \times (1+6) = 42$ propositional symbols V_{PC} ($P \in \mathcal{P}$, $C \in \mathcal{F} \cup \mathcal{P}$). Let the axiom set of T consist of formulas (1)—(3) plus the axioms

$$V_{ab1,ab1}, \tag{5}$$

$$V_{ab1,ontable}, \tag{6}$$

$$V_{ab2,ab2}, \tag{7}$$

$$V_{ab2,white}, \tag{8}$$

$$V_{ab2,red}. \tag{9}$$

Axioms (5) and (7) tell us that $ab1$ and $ab2$ are minimized. According to (6), the interpretation of the predicate constant *ontable* is allowed to vary in the process of minimizing $ab1$. This postulate is motivated by the fact that $ab1$ is introduced for the purpose of describing the locations of blocks. Axioms (8) and (9) are motivated by similar considerations: $ab2$ is used for characterizing the colors of blocks.

It can be proved that the formulas

$$ab1\ x \equiv false, \quad ab2\ x \equiv x = B \tag{10}$$

are theorems of T. These formulas, along with axioms (1) and (2), imply the desired conclusions:

$$block\ x \supset ontable\ x, \quad block\ x \wedge x \neq B \supset white\ x.$$

Remark 1. We decided for each abnormality predicate separately which predicates are varied when that particular abnormality is minimized. This is different from the use of circumscription in (McCarthy 1986), where the set of varying predicates is the same for all kinds (aspects) of abnormality. Our approach appears to make formalizations more modular.

Remark 2. The predicates V_{PC} used in this example do not have function symbols among their subscripts P, C. Our formalism allows C to be a function; this would have been essential, for instance, if we introduced the function symbol *color* instead of the predicates *white* and *red*.

5. Generating Sets

Formulas (10) provide, in a sense, a complete description of the effect of minimization in the theory T from Section 4. Let us say that a formula of a circumscriptive theory is *V-free* if it does not contain any of the symbols V_{PC}. We say that a set G of V-free theorems *generates* T, or is a *generating set* for it, if the union of G with the V-free axioms of T implies all V-free theorems of T. Using this terminology, we can say that formulas (10) generate T.

Every circumscriptive theory is generated, of course, by the set of its V-free theorems. But in this example we have a very simple generating set: a finite set of first-order (actually, even universal) formulas. Finding a simple generating set for a given circumscriptive theory is important, because the predicates V_{PC} play an auxiliary role, and we are primarily interested in V-free theorems. Methods for computing simple generating sets for some classes of circumscriptive theories based on the ideas of (Lifschitz 1985) will be presented in the final version of the paper.

6. Policy Axioms

The axioms of a circumscriptive theory which are not V-free will be called its *policy* axioms.

The policy axioms used in Section 4 tell us that some of the propositions V_{PC} are true, but say nothing about the others. We could have included the negations of any of the remaining propositions V_{PC} in the list of axioms, and that would not have changed the set of V-free theorems of T. This is a special case of the following theorem:

Proposition 2. *If V_{PC} does not occur in the axioms of a circumscriptive theory T then the circumscriptive theory obtained from T by adding $\neg V_{PC}$ to the axiom set has the same V-free theorems as T.*

Thus only "positive" information about V_{PC} is essential. We will include only such information in axiom sets.

The following notation is useful for specifying policy axioms. If $\mathcal{M} \subset \mathcal{P}$ and $\mathcal{C} \subset \mathcal{F} \cup \mathcal{P}$ then $V[\mathcal{M} : \mathcal{C}]$ stands for the conjunction $\bigwedge_{P \in \mathcal{M},\, C \in \mathcal{C}} V_{PC}$ (expressing that the predicates and functions in $\mathcal{C}$ are varied when the predicates in $\mathcal{M}$ are minimized). For instance, axioms (5)—(9) can be written in this notation as

$$V[ab1 : ab1, ontable], \tag{11}$$

$$V[ab2 : ab2, white, red], \tag{12}$$

(we drop the braces around the elements of $\mathcal{M}$ and $\mathcal{C}$).

7. Adding Axioms to a Circumscriptive Theory

The set of theorems of a circumscriptive theory T depends on the set of its axioms non-monotonically: some theorems of T may be lost if axioms are added to T. For instance, if we add the formula $\neg ontable\ B$ to the axioms of the theory from Section 4 then the first of theorems (10) will be lost (along with its corollary $ontable\ B$). In this extended theory $ab1$, like $ab2$, is equivalent to $x = B$.

There is an important special case when adding an axiom makes the set of theorems bigger, as in first-order theories. We say that a policy axiom is *pure* if it contains no symbols from $\mathcal{F} \cup \mathcal{P}$. For instance, axioms (5)—(9) are pure policy axioms.

Proposition 3. *If a circumscriptive theory T_2 is obtained from a circumscriptive theory T_1 by adding pure policy axioms then*
(i) every model of T_2 is a model of T_1,
(ii) every theorem of T_1 is a theorem of T_2.

Add, for instance, $V_{block,block}$ to the axiom set of the theory T from Section 4. The new axiom expresses our intention to minimize the predicate *block*. The new theory has some theorems that are not theorems of T, such as $block\ x \equiv x = B$. But, according to Proposition 3, no theorems of T are lost.

8. Priorities

In some cases, the axioms imply a "negative correlation" between two minimized predicates, so that one of them can be minimized only at the price of increasing the values of the other. It may be desirable to establish relative priorities between the tasks of minimizing such "conflicting" predicates (McCarthy 1986). For example, when circumscription is used for describing an inheritance hierarchy with exceptions, it may be necessary to assign a higher priority to minimizing exceptions to "more specific" default information.

In the formalism of this paper, assigning a higher priority to $P \in \mathcal{P}$ than to $Q \in \mathcal{P}$ can be expressed by the axiom $V[P : Q]$ (i.e., V_{PQ}). With this axiom, minimization guarantees that no change in the interpretation of Q would make it possible to change a value of P from *true* to *false*.

Consider, for instance, the following facts about the ability of birds to fly (McCarthy 1986). Things in general, normally, cannot fly; birds normally can. But ostriches are birds which normally cannot fly. Symbolically,

$$\neg ab1\ x \supset \neg flies\ x, \tag{13}$$

$$\neg ab2\ x \wedge bird\ x \supset flies\ x, \tag{14}$$

$$ostrich\ x \supset bird\ x, \tag{15}$$

$$\neg ab3\ x \wedge ostrich\ x \supset \neg flies\ x. \tag{16}$$

We would like to get the theorems

$$ab1\ x \equiv bird\ x \wedge \neg ostrich\ x,$$
$$ab2\ x \equiv ostrich\ x, \tag{17}$$
$$ab3\ x \equiv false.$$

Natural candidates for the policy axioms are

$$V[ab1 : ab1, flies],$$
$$V[ab2 : ab2, flies], \tag{18}$$
$$V[ab3 : ab3, flies].$$

But these axioms do not lead to the desired result, because there is a conflict between minimizing $ab2$ on the one hand and minimizing $ab1$ and $ab3$ on the other. This can be fixed by assigning different priorities to the abnormality predicates.[4] We assign the highest priority to $ab3$, because the corresponding axiom, (16), gives the "most specific" information, and the lowest priority is given to $ab1$. Formally, (18) is replaced by the following set of policy axioms:

$$V[ab1 : ab1, flies],$$
$$V[ab2 : ab1, ab2, flies], \tag{18$'$}$$
$$V[ab3 : ab1, ab2, ab3, flies].$$

This makes the theory stronger (Proposition 3). The circumscriptive theory with axioms (13)—(16) and (18′) has the desired property: it is generated by formulas (17).

9. Reasoning about Priorities

All policy axioms in the examples above are (conjunctions of) atoms. The use of more complex policy axioms allows us to formalize some forms of metalevel reasoning in the circumscriptive theory itself. Consider the following example.

Imagine that we have two sources of information about the world, and that we assume by default that any event reported by any of the sources has in fact happened:

$$\neg ab1\ x \wedge reported1\ x \supset happened\ x, \tag{19}$$

$$\neg ab2\ x \wedge reported2\ x \supset happened\ x. \tag{20}$$

Two announcements made by different sources contradict each other:

$$reported1\ A, \quad reported2\ B, \tag{21}$$

$$\neg(happened\ A \wedge happened\ B). \tag{22}$$

The circumscriptive theory with axioms (19)—(22) plus the policy axioms

$$V[ab1 : ab1, happened], \tag{23}$$

$$V[ab2 : ab2, happened], \tag{24}$$

has models of two kinds: in some of them $happened\ A$ is true, in the others $happened\ B$.

Giving a higher priority to one of the predicates $ab1$, $ab2$ would allow us to arrive at a definite conclusion about which event has actually happened. If, for instance, we consider the first source more reliable then we can add $V[ab1 : ab2]$ to the axiom set. In the extended theory we can prove $happened\ A$ and $\neg happened\ B$.

The reasoning leading to the choice of a prioritization can be formalized in the following way. Using the propositional symbols $preferred1$ and $preferred2$, we can describe our approach to establishing priorities in this example by the axioms

$$preferred1 \supset V[ab1 : ab2], \tag{25}$$

$$preferred2 \supset V[ab2 : ab1]. \tag{26}$$

In the theory with the axioms (19)—(26) we can prove

$$preferred1 \supset happened\ A, \quad preferred2 \supset happened\ B.$$

Adding the axiom $preferred1$ would make $happened\ A$ and $\neg happened\ B$ provable. In this formulation, the choice of priorities is established by logical deduction.

Acknowledgements

I am grateful to Benjamin Grosof and John McCarthy for comments and constructive criticism.

References

Lifschitz, V., Computing circumscription, *Proc. IJCAI-85* **1**, 1985, 121–127.

Lifschitz, V., Pointwise circumscription: Preliminary report, *Proc. AAAI-86* **1**, 1986, 406–410.

McCarthy, J., Programs with common sense, in *Proceedings of the Teddington Conference on the Mechanization of Thought Processes*, Her Majesty's Stationery Office, London, 1960.

McCarthy, J., Circumscription — a form of non-monotonic reasoning, *Artificial Intelligence* **13** (1980), 27–39.

McCarthy, J., Applications of circumscription to formalizing commonsense knowledge, *Artificial Intelligence* **28** (1986), 89–118.

Perlis, D., Circumscribing with sets, *Artificial Intelligence* **31** (1987), 201–211.

[4] Another approach is to use "cancellation of inheritance" axioms (McCarthy 1986).