

Circumscription

Abstract: Various people have proposed methods of modifying formal logic to describe ordinary discourse and “common sense” reasoning. One such approach is John McCarthy’s *circumscription*. There, the syntax is the same as for classical logic, but the semantics is changed; only some of the classical models of a formula are considered. We define circumscription, in two of its variants, and discuss its motivation in “negation as failure.” McCarthy hoped that his “common sense reasoning” would be computationally easier than classical logic. Unfortunately, the opposite turns out normally to be true. We state and sketch the proofs of some results. (This material will probably take 2 presentations.)

Motivation in the old Missionaries & Cannibals Problem

Problem: 3 Missionaries & 3 cannibals must cross a river, using a boat that can take only 2. Cannibals must never outnumber the missionaries on either shore, because

Constraint that McCarthy wanted to express [quoted from Wikipedia]:
Exclude conditions that are not explicitly stated. For example, the solution “go half a mile south and cross the river on the bridge” is intuitively not valid because the statement of the problem does not mention such a bridge. On the other hand, the existence of this bridge is not excluded by the statement of the problem either. That the bridge does not exist is a consequence of the implicit assumption that the statement of the problem contains everything that is relevant to its solution.

Hackneyed example:

1. *Tweety is a bird.*
2. *All not-abnormal birds fly.*

“Common sense” conclusion: *Tweety flies.*

If we later add *Tweety is a penguin* we withdraw the conclusion.

Negation as failure: We believe that Tweety is not an abnormal bird because nobody suggested that Tweety is abnormal.

More History, *again quoted from Wikipedia* **Circumscription** was later used by McCarthy to formalize the implicit assumption of inertia: things do not change unless otherwise specified. Circumscription seemed to be useful to avoid specifying that conditions are not changed by all actions except those explicitly known to change them; this is known as the frame problem. However, the solution proposed by McCarthy was later shown leading to wrong results in some cases,

Formal Definition:

- We are given a first order formula ϕ
- and a structure $\mathfrak{A} = \langle A; R_1^{\mathfrak{A}}, R_2^{\mathfrak{A}}, \dots \rangle$ for the language of ϕ .
- Now consider *all* models $\mathfrak{A}' = \langle A; R_1^{\mathfrak{A}'}, R_2^{\mathfrak{A}'}, \dots \rangle$ of ϕ with the same universe A .
- The older **Closed World Assumption** (CWA) [Reiter]:
Each $R_i^{\mathfrak{A}}$ should be true, on each individual tuple of objects, if and only if every $R_i^{\mathfrak{A}'}$ is also true.

There are obvious “problems” with CWA:

For example, for

$$\phi = (\text{takes}(\text{calculus}, \text{john}) \vee \text{takes}(\text{gothicArt}, \text{john})),$$

both $\text{takes}(\text{calculus}, \text{john})$ and $\text{takes}(\text{gothicArt}, \text{john})$ would be inferred to be false.

- **Circumscription** Think of adding a name c_a for each element $a \in A$.

Circumscription requires that **the set of atomic statements $R_i(c_{a_1}, \dots, c_{a_m})$ true in \mathcal{A} be *minimal*** — that is, there is no way to get a model of ϕ with universe A and a proper subset of those statements true.

Now, in the **Example above**, with universe $\{john, xizhong, ryan, calculus, gothicArt, modernFudge\}$, we would get 2 minimal interpretations:

1. $takes(calculus, john)$
— and $takes$ is false of all other combinations, and
2. $takes(gothicArt, john)$.

In particular, we could infer

$$\neg(takes(calculus, john) \wedge takes(gothicArt, john)).$$

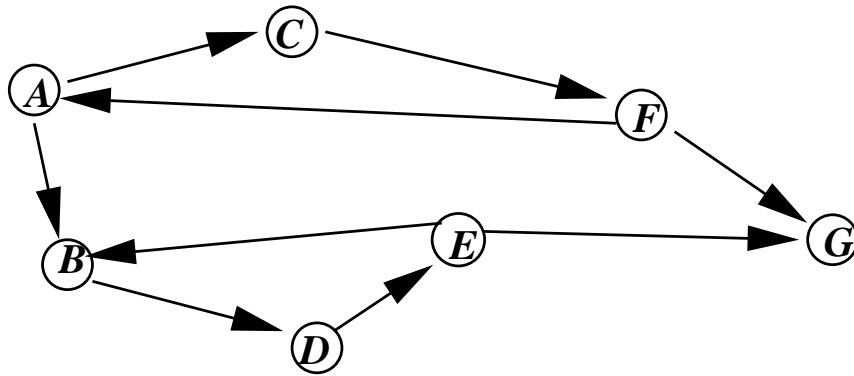
We'd also, nicely, infer

$$\neg takes(john, modernFudge).$$

A standard example: When we give a database, we represent the arguments for which the relations are false.

We **assume** all other atomic statements are false.

Another standard example: *Transitive closures*:



$edge(A, B)$

$edge(A, C)$

$edge(B, D)$

$edge(C, F)$

$edge(D, E)$

$edge(E, B)$

$edge(E, G)$

$edge(F, A)$

$edge(F, G)$

$edge(X, Y) \rightarrow path(X, Y)$

$path(X, Y) \wedge path(Y, Z) \rightarrow path(X, Z)$

Notes:

1. We didn't say where $\neg edge$ or $\neg path$ is true.
2. We *cannot* say, in *classical first order logic*, when $\neg path$ is true.

- Lifschitz created several variants.

In one, some of the relation symbols are put into 2 separate groups:

Fixed relations: When we compare 2 structures \mathfrak{A} and \mathfrak{A}' , we require that the 2 structures interpret the fixed relations the same way.

Varying relations: We allow changing the interpretations of these relations between structures \mathfrak{A} and \mathfrak{A}' , but in minimizing the true atomic formulas $R_i(c_{a_1}, \dots, c_{a_n})$, we ignore the varying relations.

Tweety Again:

1. *Tweety is a bird.*
2. *All not abnormal-birds fly.*

Treat *fly* as a varying relation.

Minimize the interpretations of relations *bird* and *abnormal-bird*.

- We said *Tweety* is a *bird*.
- We did *not* say that *Tweety* is an *abnormal bird*.
- So we infer that *Tweety flies*.

Note that the “All not abnormal-birds fly” is basically a formulation of a stereotype.

I think the argument here is that, although stereotypes can be abused, they are also a necessary part of our thinking process. For example, we can't always take the time to consider carefully whether the building we're is ready to fall down. over.

Circumscription does not seem strong enough to explain McDermott’s “Yale Shootout Problem.” The problem comes because “common sense” seems to assign implicit priorities on minimization:

I’ll give a simpler example, a standard game example. **2 players take turns making moves. The winner is the one who moves last — the one who moves so that his opponent to has no legal move.**

$$\begin{array}{ll}
 \text{move}(A, B) & \text{move}(A, C) \\
 \text{move}(B, D) & \text{move}(B, E) \\
 \text{move}(C, E) & \text{move}(C, F) \\
 \text{move}(D, F) & \text{move}(D, G) \\
 & \vdots \\
 \text{move}(X, Y) \wedge \neg \text{winning}(Y) & \rightarrow \text{winning}(X)
 \end{array}$$

Simple example where circumscription doesn’t capture “the intuition” of those assertions:

$$A \xrightarrow{\text{move}} B \xrightarrow{\text{move}} C$$

By popular demand: *Yale shootout problem* (modified):

$$\begin{array}{ll}
 \text{loaded}(t) \wedge \text{shoots}(t) & \rightarrow \text{noise}(t) \\
 & \text{loaded}(0) \\
 \text{loaded}(s) \wedge \neg \text{shoots}(s) \wedge \text{succ}(s, t) & \rightarrow \text{loaded}(t) \\
 \text{triggers}(t) & \rightarrow \text{shoots}(t) \\
 & \text{triggers}(1) \\
 & \text{succ}(0, 1)
 \end{array}$$

McCarthy hoped that computation with circumscription would be *easier* than for classical logic. **BUT**

Theorem [Martin Davis]: *There is no axiom system for circumscriptive reasoning.*

Indeed, consider the following problem: Given formulas ϕ, ψ , does ψ follow from the circumscription of ϕ ?

The set of such ψ 's is neither r.e. nor co-r.e.

Theorem [Cadoli]: (In obvious propositional logic analogue) *Determining whether a truth assignment is a minimal model of a propositional formula is NP-complete.*

Theorem [Eiter & Gottlob]: *Inference in propositional circumscription is Π_2^P complete.*

Theorem [JSS]: *Even when no varying predicates are allowed, determining whether a first order formula ϕ has a countable minimal model is Σ_2^1 -complete over arithmetic.*

Difficulty: Every finite partial ordering has minimal elements, but not every infinite one does.

Corollary: *Determining whether a first order formula ψ follows from a first order formula ϕ under circumscription, limited to countable models, is Π_2^1 -hard over arithmetic.*

Theorem [Eiter & Gottlob]: *Inference in circumscription (with first order formulas), where no cardinality assumptions are made, is the same difficulty as validity in full second order logic.*

Peano's axioms (approx., + a speck of 2nd order logic)

- 1st order: quantify over (all) *elements* of the structure.
- 2nd order: quantify over *both* elements and (all) *subsets* and *relations* on the structure.
- A standard example of second order logic, describing the natural numbers $\mathfrak{N} = \langle \mathbb{N}; 0, 1, succ, +, \cdot, < \rangle$ [basically Peano]

$$\mathbf{P} = \{ \begin{array}{l} \forall x \forall y \forall z (x < y \wedge y < z \rightarrow x < z) \\ \forall x \forall y (x < y \vee x = y \vee y < x) \\ \forall x (x \neq 0) \quad \forall x (succ(x) \neq x) \\ \forall x (x = 0 \vee \exists y (x = succ(y))) \\ \forall x \forall y (succ(x) = succ(y) \rightarrow x = y) \\ \forall x \forall y (x < succ(y) \leftrightarrow (x < y \vee x = y)) \\ \forall x (x + 0 = x) \quad \forall x (x \cdot 0 = 0) \\ \forall x \forall y (x + succ(y) = succ(x + y)) \\ \forall x (x \cdot succ(y) = x \cdot y + x) \\ \\ \forall N (N(0) \wedge \forall x (N(x) \rightarrow N(succ(x))) \rightarrow (\forall x (N(x)))) \end{array} \}$$

Call first order axioms above \mathbf{P}^f .

Davis' complexity result: essentially from \mathbf{P}^f .

For any first formula ϕ of this language, form ϕ^N from ϕ by

- replacing each subformula $\forall x \psi$ with $\forall x (N(x) \rightarrow \psi)$, and
- replacing each subformula $\exists x \psi$ with $\exists x (N(x) \wedge \psi)$.

Let $\delta = \bigwedge \mathbf{P}^f \wedge \forall x (N(x) \rightarrow N(succ(x)))$. Minimize just N :

For any sentence ϕ of number theory,

$$\delta \models_{Circ} \phi^N \text{ iff } \mathfrak{N} \models \phi.$$

What does a model \mathfrak{M} of \mathbf{P}^f look like?

1. They are linearly ordered by $<$, with least element 0.
2. Each element x has a $<$ -successor, $\text{succ}(x)$.

So \mathfrak{M} starts out

$$0 < \text{succ}(0) < \text{succ}^2(0) < \text{succ}^3(0) < \dots$$

called the *standard part* of \mathfrak{M} . *(Sometimes it's all of \mathfrak{M} .)*

(A model of \mathbf{P}^f is called *standard* if all its elements are in its standard part. We can show that all standard models are isomorphic to the actual natural numbers.)

3. Each element x except 0 has a $<$ -predecessor, $\text{succ}^{-1}(x)$.

So the non-standard part of \mathfrak{M} has no least element.

So, what does a model of δ look like?

1. From (1,2) above, the standard part of \mathfrak{M} must \subseteq any relation N satisfying δ .

But interpreting N as the standard part satisfies δ ,

so the standard part is the minimum interpretation of N ,

giving Davis' result.

Another example: Here minimize both N and R :

$$\begin{aligned}\sigma &= \delta \wedge \forall x(N(x) \rightarrow R(x)) \\ &\wedge (\forall x(R(x)) \vee \exists y \forall x(R(x) \leftrightarrow x < y)).\end{aligned}$$

- If \mathfrak{M} is standard, the minimum (and only) interpretation has N and R both equal to the standard part too.
- Otherwise, we'll show \mathfrak{M} *can't be minimal*: For suppose $\mathfrak{M} = \langle M; 0, 1, +, \cdot, succ, N^{\mathfrak{M}}, R^{\mathfrak{M}} \rangle \models \sigma$ is a minimal model.

Case 1: $N^{\mathfrak{M}}$ is not just the standard part of \mathfrak{M} :

If we shrink the interpretation of N to the standard part of \mathfrak{M} , we'll still (obviously) have all the axioms of σ satisfied, so \mathfrak{M} isn't minimal after all.

Case 2: (i) \mathfrak{M} is not a standard model (i.e., it has some non-standard elements), (ii) $N^{\mathfrak{M}}$ is just the standard part of M , and (iii) $R^{\mathfrak{M}}$ is all of \mathfrak{M} .

Then we can find a smaller model by choosing any non-standard y and interpreting R to be y 's predecessors. So \mathfrak{M} isn't minimal.

Case 3: (i) \mathfrak{M} is not a standard model (i.e., it has some non-standard elements), (ii) $N^{\mathfrak{M}}$ is just the standard part of \mathfrak{M} , and (iii) $R^{\mathfrak{M}}$ is the set of all predecessors of some element y .

Since $y \notin R$, y is nonstandard. Since there is no least nonstandard element, there is a smaller nonstandard element y' .

And if we interpret R to be the set of predecessors of y' , we'll get a smaller interpretation of R which still (obviously) satisfies σ . So \mathfrak{M} isn't minimal.

- So circumscription essentially *forces* \mathfrak{M} to be isomorphic to the natural numbers.

Hierarchy of 2nd order formulas:

- A Δ_0^1 formula is a first order formula has no quantifiers over sets or relations — although it may have set/relation variables.
- A Σ_1^1 formula is of the form $\exists X_1 \exists X_2 \cdots \exists X_n \phi$ where ϕ is Δ_0^1 .
- A Π_1^1 is of the form $\forall X_1 \forall X_2 \cdots \forall X_n \phi$ where ϕ is Δ_0^1 .
- A Σ_2^1 formula is of the form $\exists X_1 \exists X_2 \cdots \exists X_n \phi$ where ϕ is Π_1^1 .
- A Π_2^1 is of the form $\forall X_1 \forall X_2 \cdots \forall X_n \phi$ where ϕ is Σ_1^1 .
- \vdots

Proposition: *Saying that $\phi \models_{\text{Circ, cntbl mdl's}} \psi$ is Π_2^1 definable over \mathfrak{N} .*

(I just coined that symbol $\models_{\text{Circ, cntbl mdl's}}$; I think its meaning is obvious.)

Proof sketch: It's equivalent to show that

Saying that ϕ has a countable minimal model is Σ_2^1 definable over \mathfrak{N} .

- Consider a formula $\phi(N, R, S, T)$ with unary relations N, R to be minimized and unary relations S, T to be fixed.
- Since we're looking at countable models, we may assume, w.l.o.g., that their universes are sets of natural numbers.

We'll use 2nd order variable U to identify their universes.

- So we can write (the de-abbreviation of)

$$\begin{aligned} \exists U \exists N \exists R \exists S \exists T \quad & (\quad N, R, S, T \subseteq U \wedge \phi^U(N, R, S, T) \\ & \wedge \forall N' \subseteq N \forall R' \subseteq R \\ & \quad (\phi^U(N', R', S, T) \rightarrow R = R' \wedge U = U') \\ &) \end{aligned}$$

which can be shown to be equivalent to a Σ_2^1 formula by familiar quantifier manipulations.