Logic(s) and Artificial Intelligence

Jacek Malec
Department of Computer Science
Lund University
Box 118
Ole Römers väg 3
221 00 LUND, Sweden
email: jacek.malec@cs.lth.se

March 25, 2002

Plan for this hour:

- Background
  - Logic
  - Artificial Intelligence
- First Order Predicate Calculus (Resolution, Horn clauses, Prolog)
- Reasoning about action and change (non-monotonicity)
- Modal logics
- Rationality of an agent (BDI models)
- Other tracks of research
- The future

Background

- Logic
  - Correct reasoning (Aristotle)
  - Language of mathematics (Hilbert)
  - Semantics (Tarski)
  - Incompleteness (Gödel)
  - Other (non-classical) approaches
- Artificial Intelligence
  - 1956: Dartmouth (McCarthy, Minsky, Newell, Simon)
  - Principle of rationality
  - Physical symbol system hypothesis
  - Situated agents

A *physical symbol system* consists of a set of entities, called symbols, which are physical patterns that can occur as components of another type of entity called *an expression* (or symbol structure). Thus, a symbol structure is composed of a number of instances (or tokens) of symbols related in some physical way (such as one token being next to another). At any instant of time the system will contain a collection of these symbol structures. Besides these structures, the system also contains a collection of processes that operate on expressions to produce other expressions: processes of creation, modification, reproduction and destruction. A physical symbol system is a machine that produces through time an evolving collection of symbol structures. Such a system exists in a world of objects wider than just these symbolic expressions themselves.

The physical symbol system hypothesis.

A physical symbol system has the necessary and sufficient means for general intelligent action.

Roughly every fifty years humanity solves a great mystery of science. We have a chance to solve such a mystery now: **How does the human mind work?** There are many corollary questions:

- Where does the mind reside?
- What is the nature of memory?
- What are the roles of emotions?
- What representations does the brain use?
- What does our visual system compute?
- How did evolution shape us?
- How do we learn?
- What is consciousness?

Keys to intelligence: self-adapting perceptual systems, motor systems and language-related modules (in contrast to reasoning, planning and knowledge representation)

[Rodney Brooks, MIT AI Lab director]

First Order Logic

- Syntax
  - Predicates (relations, properties)
  - Constants, variables
  - Functions
  - Terms
  - Atomic sentences
  - Well-formed formulae
- Semantics
  - interpretations
  - models
  - satisfiablity
  - validity
  - inconsistency

Ex.

$At(Robot, PostRoom)$

$\neg Empty(JaceksMailbox)$

PICKPOST:

Preconditions:

$At(Robot, PostRoom)$

Deletions:

$\neg Empty(JaceksMailbox)$

Additions:

$Empty(JaceksMailbox)$

$Holding(JaceksMail)$

First Order Logic

- A formal system:

  axioms + rules of inference
- Theoremhood, provability relation
- Syntax vs. semantics ($\vdash$ vs $\models$)
  - completeness (every valid formula has a proof)
  - soundness (every theorem is valid)
  - (semi)decidability (there exists a procedure to decide whether a formula is valid)

A first order system

Logical axioms

A1 $A \to (B \to A)$

A2 $(A \to (B \to C)) \to ((A \to B) \to (A \to C))$

A3 $(\neg B \to \neg A) \to ((\neg B \to A) \to B)$

A4 $(\forall x_i)A(x_i) \to A(t)$

A5 $(\forall x_i)(A \to B) \to (A \to (\forall x_i)B)$

Theory axioms (e.g. partial order)

T1 $(\forall x)(\neg(x < x))$

T2 $(\forall x, y, z)(x < y \wedge y < z \to x < z)$

Rules of inference

- Modus Ponens
- Generalisation

Resolution: a generalization of Modus Ponens (1965, J. Robinson)

$$\frac{\begin{array}{c}A_1 \vee A_2 \vee \ldots \vee A_m \vee \neg C\\ B_1 \vee B_2 \vee \ldots \vee B_n \vee C\end{array}}{A_1 \vee A_2 \vee \ldots \vee A_m \vee B_1 \vee B_2 \vee \ldots \vee B_n}$$

Modus Ponens:

$$\frac{\begin{array}{c}\neg P \quad \vee Q\\ P\end{array}}{Q}$$

Universal Instantiation:

$$\frac{\forall x A}{A'(x \to t)}$$

Formal System for FOPC:

language of FOPC, axioms + RES and UI

- Horn clauses (a restricted FOPC)
- Prolog

Reasoning about action and change

- Situation calculus
- Closed World Assumption and relatives
- Negation as failure (PROLOG)
- Default reasoning
- Circumscription

Situation calculus

a situation — frame, state of the world

$At(Robot, PostRoom)$

$$At(Robot, PostRoom, S_0) \wedge$$
$$\wedge \neg Empty(JaceksMailbox, S_0)$$

$Result(PickPost, S_0) = S_1$

$Result(GoToJaceksRoom, S_1) = S_2$

$$(\forall s)(\neg Empty(JaceksMailbox, s) \rightarrow$$
$$\rightarrow Holding(JaceksMail, Result(PickPost, s))$$

Frame problem

$$(\forall a, s, x) Holding(x, s) \wedge a \neq Release \rightarrow$$
$$\rightarrow Holding(x, Result(a, s))$$

$$(\forall a, s, x) \neg Holding(x, s) \wedge$$
$$\wedge (a \neq PickPost \vee Empty(JaceksMailbox, s)) \rightarrow$$
$$\rightarrow \neg Holding(x, Result(a, s))$$

Assume

$\neg Empty(EricsMailbox, S_0)$

What happens with Eric's mailbox after the one of Jacek has been emptied?

We would need thousands of axioms!

Closed World Assumption:

Things that cannot be proven true are probably false

Consider:

$\Delta =$

$$\{F(Stockholm), F(Malmoe), F(Copenhagen), F(Oslo), Conn(Oslo, Copenhagen), Conn(Stockholm, Oslo), \forall (x, y, z)(Conn(x, y) \wedge Conn(y, z) \rightarrow Conn(x, z))\}$$

Because it's not the case that $Connect(Malmoe, Stockholm)$, CWA gives us immediately $\neg Connect(Malmoe, Stockholm)$.

But let us add $Connect(Malmoe, Copenhagen)$ to Δ. $Connect(Malmoe, Stockholm)$ can be proven now, so $\neg Connect(Malmoe, Stockholm)$ does not appear as a consequence.

Non-monotonicity!

CWA is syntax-dependent:

If $\Delta =$
$\{Single(John), Single(Mary), Clever(Kent)\}$

then CWA gives us:
$\{\neg Clever(John), \neg Clever(Mary), \neg Single(Kent)\}$.

But if $\Delta =$
$\{\neg Married(John), \neg Married(Mary), Clever(Kent)\}$

then CWA gives us:
$\{\neg Clever(John), \neg Clever(Mary), \neg Married(Kent)\}$.

Domain Closure Assumption:

The only objects in the domain are those which can be named by the object constants and functions available in the language.

Unique-Names Assumption:

If we cannot prove that two ground atoms are equal, then we can assume that they are different.

Default logic (Reiter)

default rules:

$$\frac{\alpha(x)|\beta(x)}{\gamma(x)}$$

If $\alpha(x)$ and it is consistent to assume $\beta(x)$ then infer $\gamma(x)$.

Example:

$$\frac{Bird(x)|Flies(x)}{Flies(x)}$$

Default rules express knowledge about what uses to be true. E.g. the default above says: Birds usually fly.

A consistency-based approach

Preferential semantics

Ex. language p, q

Models: $I_{\neg p,\neg q}, I_{\neg p,q}, I_{p,\neg q}, I_{p,q}$

$<$ - preference among models

$I_{\neg p,\neg q} < I_{\neg p,q}$, $I_{\neg p,\neg q} < I_{p,q}$, $I_{\neg p,q} < I_{p,\neg q}$, $I_{p,q} < I_{p,\neg q}$.

$\emptyset$ has all four Is as models. But we prefer $I_{\neg p,\neg q}$! Thus $\emptyset \models \neg p$.

$\{q\}$ has two (both minimal) models. Observe that $\neg p$ does not hold in one of them, thus cannot be inferred! Non-monotonicity!

Circumscription

Essentially: an additional second order axiom:

$$A(P;Q) \wedge (\forall p, q_1, q_2, ..., q_m)(A(p, q_1, ..., q_m) \wedge$$

$$\wedge((\forall x)(p(x) \rightarrow P(x)) \rightarrow (\forall x)(P(x) \rightarrow p(x)))$$

where $A(P;Q)$ is a formula with a predicate symbol P and m predicate symbols Q_i (s.c. variable predicates).

The effect of circumscribing P in a theory A is to limit the instances that satisfy P to only those that are necessary in light of A.

Typical use:

$$Frog(x) \wedge \neg Ab(x) \rightarrow Green(x)$$

$$Grassfrog(x) \rightarrow Frog(x)$$

$$Grassfrog(x) \rightarrow Brown(x)$$

$$Brown(x) \rightarrow Ab(x)$$

$$Frog(Kermit), Grassfrog(Scooter)$$

Minimising $Ab(x)$ yields intended results.

Modal logics

- standard systems, semantics
- knowledge and belief
- temporal logics

Modal Logic

Syntax

- the language of FOPC
- $\Box F$ (LF) necessity
- $\Diamond F$ (MF) possibility

Semantics (Kripke)

A modal frame M is a structure $< W, D, R, F >$ where

- W is a non-empty set (of possible worlds);
- D is a non-empty domain (of individuals);
- R is a binary relation (of accessibility) on W;
- F is an interpretation function for function and relation symbols, parametrized by $w \in W$.

$$M \models_{g,w} A$$

interpretation g satisfies the wff A at the world w in the frame M.

Properties of relation R determines the validity in a modal logic.

- reflexive - system T;
- reflexive and transitive - system S4;
- transitive and Euclidean - system S45 (weak S5);
- equivalence - system S5.

Formal system:

formal system of FOPC, and

R $\frac{A}{\Box A}$ (necessitation);

K $\Box(A \to B) \to (\Box A \to \Box B)$;

D $\Box A \to \Diamond A$;

T $\Box A \to A$;

4 $\Box A \to \Box\Box A$;

5 $\Diamond A \to \Box\Diamond A$.

- KT is T
- KT4 is S4
- KD45 is S45
- KT5 is S5

Hintikka - epistemic logic ($\Box$ means “knows”)

S5 - logic of knowledge

S45 - logic of belief

Ex.

$$(K_{me}K_{you}key(me) \wedge K_{me}send(me, you, msg)) \to$$

$$\to FK_{you}contents(msg)$$

i.e.

> If I know that you know my public key and I know that I have sent you an (encrypted) message then sometime in the future you will know the contents of that message.

Temporal logics

Syntax:

- the language of FOPC
- FA - sometime in the future;
- PA - sometime in the past;
- GA - always in the future;
- HA - always in the past.

Transition relation R models temporal precedence.

Some (interesting) systems introduce also:

- $\bigcirc A$ (Next A)
- $A\mathcal{U}B$ (A until B)

Axiomatisation of (propositional) TL(s)

1. Theorems of propositional calculus
2. $G(A \to B) \to (GA \to GB)$
3. $H(A \to B) \to (HA \to HB)$
4. $A \to HFA$
5. $A \to GPA$
6. GA if A is an axiom
7. HA if A is an axiom
8. $FFA \to FA$
9. $(PA \to PB) \to$ $(P(A \wedge B) \vee (P(A \wedge PB) \vee P(PA \wedge B)))$
10. $(FA \to FB) \to$ $(F(A \wedge B) \vee (F(A \wedge FB) \vee F(FA \wedge B)))$
11. $GA \to FA$
12. $HA \to PA$
13. $FA \to FFA$

Logic on temporal intervals (Allen)

$i1, i2$

- $during(i1, i2)$
- $before(i1, i2)$
- $overlap(i1, i2)$
- $meets(i1, i2)$
- $equal(i1, i2)$

Reification:

$$Holds(A, i)$$

$$Occur(e, i)$$

Rationality of an agent (BDI models)

- specification
- implementation
- verification

Bratman

Georgeff, Rao, Singh

Fisher, Wooldridge

BDI - belief, desire, intention

- Bel - system S4 or S45, accessibility relation B
- Des - system KD, accessibility relation D
- Int - system KD, accessibility relation I
- (sometimes) K_h (know-how) - system S5

plus a temporal logic (linear or branching).

Often CTL* (computational tree logic, branching discrete time)

CTL* has two types of formulae: state f. and path f. Temporal operators: $\bigcirc$ and $\mathcal{U}$. Path operators: **A** and **E**. Other temporal modalities may be defined.

If A is a CTL* formula, then $Bel(A)$, $Des(A)$ and $Int(A)$ are BDI-logic formulae. Often $Bel(i, A)$, $Des(i, A)$ and $Int(i, A)$.

A simple example:

Attack is the best defence.

$$Bel(me, F(Des(you, attack(you, me)))) \rightarrow$$

$$\rightarrow Int(me, \bigcirc attack(me, you))$$

i.e.

> If I believe that you desire to attack me, then I intend to attack you at the next time moment.

BDI logics can express:

- dynamism (via temporal component);
- information (via modal belief component, possibly incorrect);
- motivation
  - desires represent goals that the agent must eventually satisfy;
  - intentions represent the goals that the agent is *actively* trying to satisfy.

The interaction between the components gives us the expressive power, but makes the system (almost) intractable.

Some properties (axioms)

- Satisfiability: $Int(i, P) \rightarrow \mathbf{E}FP$
- Temporal consistency: $(Int(i, P) \wedge Int(i, Q)) \rightarrow Int(i, FP \wedge FQ)$
- Persistence does not entail success: $\mathbf{E}G(Int(i, P) \wedge \neg P)$

Model checking algorithms exist!

Proof systems exist!

Other tracks of research

- description logics
- multi-valued logics
- fuzzy logic
- Labeled Deduction Systems
- approximation
- ...

Description logics (Baader)

inheritance networks, semantic networks, frames

- concepts (unary predicates)
- roles (binary predicates)

$C_1 \sqcap \ldots \sqcap C_n$	and
$C_1 \sqcup \ldots \sqcup C_n$	or
$\neg C$	not
$\exists R.C$	some R C
$\forall R.C$	all R C
$(\geq nR)$	atleast n R
$(\leq nR)$	atmost n R
$R_1 \sqcap \ldots \sqcap R_n$	and

$Frog \doteq Animal \sqcap \forall colour.Green$

DL belong to the *guarded fragment* of FOPL.

Multi-valued logics

- 3-valued (Kleene, Łukasiewicz, Sandewall and Doherty)

  true, false, undecided
- 4-valued logic (Belnap)

  true, false, unknown, contradictory
- multi-valued logic
- fuzzy logic (Zadeh)

Would be good to have (JM):

- Effective algorithms for reasoning about action and change;
- Reasoning with crisp (hard-real-time) deadlines;
- Tools for automated reasoning.

Logic and (symbolic) Artificial Intelligence inspire each other in many different ways.