WHAT ARTIFICIAL INTELLIGENCE NEEDS FROM SYMBOLIC LOGIC

John McCarthy, Stanford University
mccarthy@stanford.edu
http://www-formal.stanford.edu/jmc/
December 4, 2006

The goal of artificial intelligence research is human-level AI. Logical AI is an approach. It requires mathematical logic with human-level expressiveness both in the formulas it can include and in the reasoning steps it allows.

Here are the topics.

- What is logical AI?
- The common sense informatic situation
- Relevant history of logic

- Problems with logical AI
- Nonmonotonic reasoning
- Domain dependent control of reasoning
- Concepts as objects
- Contexts as objects
- Partially defined objects
- Self-awareness
- Remarks and references

LOGICAL AI

- Logical AI proposes computer systems that represent what they know about the world by sentences in a suitable mathematical logical language. It achieves goals by *inferring* that a certain strategy of action is appropriate to achieve the goal. It then carries out that strategy, using observations also expressed as logical sentences.

- If *inferring* is taken as *deducing* in (say) first order logic, and the logical language is taken as a first order language, this is inadequate for human-level AI. Extended inference mechanisms are certainly required, and extended languages are probably required.

- Much AI research and all practical applications today have more modest goals than human-level AI. Many of them use weaker logics that seem more computationally tractable.

- A rival approach to AI is based on imitating the neural structures of human and animal intelligence. So far the ability to understand and imitate these structures has been inadequate.

THE RELEVANT DEVELOPMENTS IN MATHEMATICAL LOGIC

• Leibniz, Boole, Frege, and Peirce all expected that mathematical logic would apply to human affairs. Leibniz was explicit about replacing argument by calculation. We take this as implying that common sense reasoning would be covered by formal logic.

• Leibniz's goal for logic might be described as *human level expressiveness*. We'll see why it didn't work, but we hope it can be made to work.

• Frege gave us first order logic, which Gödel proved complete. No proper extension of first order logic is true in all interpretations. We'll see that nonmonotonic extensions can be made that are true in *preferred interpretations*.

• Whitehead and Russell's *Principia Mathematica* was an unsuccessfull start on a practical system.

• Gödel's arithmetization of metamathematics was a step towards human-level expressiveness. More readable and computable representations are needed for computation. I recommend Lisp, but a more expressive abstract syntax that provided for expressions with bound variables would be better.

• What about incompleteness? No time to say more than humans are also incomplete.

• In principle, set theory (ZFC) is an adequately expressive language for mathematics—and for AI also.

• However, all these logical systems are monotonic.

NONMONOTONIC REASONING

Pre-1980 logical systems are almost all monotonic in the sense that if A is a set of sentences such that $A \vdash p$ and $A \subset B$, then $B \vdash p$. Likewise for $A \models p$.

• Leibniz, Boole and Frege all expected that mathematical logic would reduce argument to calculation. A major reason why Leibniz's hope hasn't been realized is the lack of formalized nonmonotonic reasoning. Unfortunately, it's not the only reason.

• We concentrate on one form of nonmonotonic reasoning—finding *the minimal models* according to some ordering $<$ of interpretations. Let $A(P; Z; C)$ be an axiom involving a vector P of predicate and function symbols and two other vectors Z and C. We minimize P according to the ordering, letting Z vary and holding the predicates C constant. The formula is

$$A(P : Z; C) \wedge (\forall P' \; Z')(A(P', Z'; C) \rightarrow \neg(P' < P)). \tag{1}$$

• The important special case is *circumsription* in which the ordering relation is

$$P \leq P' \equiv (\forall x)(P(x) \rightarrow P'(x)). \tag{2}$$

• A simple class of circumscriptive theories minimizes an abnormality predicate ab.

• A simple theory of which objects fly has the axiom

FLY1:

$$
\begin{array}{l}
\neg ab(aspect1(x)) \rightarrow \neg flies(x) \\
\qquad \text{Objects normally don't fly} \\
bird(x) \rightarrow ab(aspect1(x)) \\
bird(x) \wedge \neg ab(aspect2(x)) \rightarrow flies(x) \\
ostrich(x) \rightarrow bird(x) \\
ostrich(x) \rightarrow ab(aspect2(x)) \\
ostrich(x) \wedge \neg ab(aspect3(x) \rightarrow \neg flies(x)
\end{array}
\tag{3}
$$

If we circumscribe the predicate ab in the axiom FLY1, varying the predicate flies and holding $bird$ and $ostrich$ constant, we will conclude that those objects that fly are the birds that are not ostriches.

We can elaborate the FLY1 theory by conjoining additional assertions before we circumscribe ab. For example,

moreFLY:

$$
\begin{array}{l}
bat(x) \rightarrow ab(aspect1(x)) \\
bat(x) \wedge \neg ab(aspect4(x)) \rightarrow flies(x) \\
penguin(x) \rightarrow bird(x) \\
penguin(x) \rightarrow ab(aspect2(x)) \\
penguin(x) \wedge \neg ab(aspect3(x) \rightarrow \neg flies(x).
\end{array} \tag{4}
$$

The circumscription then gives that the flying objects consist of bats and the birds that are neither ostriches nor penguins. Unfortunately, simple abnormality theories are insufficient for formalizing common sense and more elaborate nonmmonotonic reasoning is needed.

THE COMMON SENSE INFORMATIC SITUATION

• Reaching human-level expressiveness requires logical language that can express what humans do in the common sense informatic situation.

• A theory T used by the agent is open to extension to a theory T' by adding facts taking into account more phenomena.

• The objects and other entities under consideration are incompletely known and are not fully characterized by what is known about them.

• Most of the entities considered are never fully defined.

• The informatic situation itself is an object about which facts are known. This human capability is not used in most human reasoning, and very likely animals don't have it.

• Many of the objects considered are examples of *natural kinds* which can be identified by simple criteria in common situaations but about which there is more to be learned. Example: A child learns to identify lemons in the store as small yellow fruit, but lemons share a complex biology. It helps the child that the store does not have a continuum of fruits between lemons and oranges.

• The thinking doable in logic is connected with lower level mental activity. Consider getting car keys from ones pocket.

• Science, mathematics, and logic are imbedded in common sense. That's why articles and books on these subjects have words in addition to the formulas.

THE CSIS IN MATHEMATICS

> "The development of mathematics toward greater precision has led, as is well known, to the formalization of large parts of it, so that one can prove any theorem using nothing but a few mechanical rules."

This is the first sentence of Gödel's 1931 paper on incompleteness. It illustrates that mathematics is done within the common sense informatic situation.

- Consider the phrases "toward greater precision", "as is well known", and "mechanical rules".
- The first two are inherently imprecise, but Gödel is not to be faulted for using them.
- "mechanical rules" was imprecise in 1931,

but Gödel later considered that Turing had made it precise.

• Human-level expressiveness requires such terms. In logic they must be treated with weak axioms, i.e. giving up hope of if-and-only-if definitions. But there has to be more.

> "Note that the class A in Axiom B1 and the class B in Axioms B5-B8 are not fully defined, since nothing is said about those sets which are not pairs (triples), whether or not they belong to A (B).", p. 37, vol. II.

The second quotation is directly metamathematical, giving advice to the reader not expressible in the theory being developed.

INDIVIDUAL CONCEPTS AND PROPOSITIONS AS OBJECTS

• Since individual concepts and propositions can be discussed as objects in natural language, they probably must also be objects in a logical language useful for human level AI.

• $Knows(Pat, TTelephone(MMike))$ is how I say that Pat knows Mike's telephone number. $Dials(Pat, Telephone(Mike))$ is the action of Pat dialing the number. Thus $MMike$ is the concept of Mike.

$$\bullet (\forall x)(Puppy(x, Lassie) \rightarrow Knows_{dog}(Lassie, LLocation_{dog}(CConcept_{dog}(x)))) \tag{5}$$

is how we say that Lassie knows the location of all her puppies. $CConcept_{dog}(x)$ is a dog's concept of the object x, very likely different from a human's concept.

• The AI programs of Stuart Shapiro and Len Schubert use concepts as objects.

CONTEXTS AS OBJECTS

• Informal human reasoning always operates within a context but can switch from one context to another and can relate entities belonging to diffierent contexts.

• Our candidate for human-level expressiveness is to make contexts into logical objects and to include in our logical language relations among contexts and relations among the values of expresssions in different contexts.

• Our examples take the form of context: expression.

• [These slides: I = John McCarthy], [Sherlock Holmes stories: Detective(Holmes)], [Literary history: Sherlock Holmes was named after Oliver Wendell Holmes Sr., whom Conan Doyle admired as a medical detective.]

• Formal example: The contexts for an Air Force database and a General Electric database give different prices for the jet engine GE721. The AF database price includes a spare parts kit and GE doesn't.

$$\begin{array}{l}\text{Outer context :}\\ (AFdbase : Price(GE721) = \\ \quad GEdbase : (Price(GE721)) \\ \quad +GEdbase : (Price(\text{spare-parts-kit}(GE721)))\end{array} \tag{6}$$

• Reasoning in limited contexts may serve to isolate from each other contexts whose "truths" are mutually inconsistent.

PARTIALLY DEFINED OBJECTS

• Exactly what ice and rocks constitute Mount Everest is not definite. It is definite that the mountain was climbed in 1953.

• Exactly what constitutes the wants of the United States is not definite. It is definite that the United States wanted Iraq to withdraw from Kuwait in 1990.

• We can deal with partially defined objects by giving weak axioms—i.e. not requiring necessary and sufficient conditions.

• The semantics of theories with partially defined objects seems obscure to me.

• Slogan: Build solid theoretical structures on foundations of semantic quicksand. Euclid did that.

SELF-AWARENESS

- A human can be aware of his intentions, hopes, fears, knowledge of a domain, non-knowledge. A language with human-level expressiveness will have terms for these.

- Self-awareness is a recent evolutionary development, and is only partial in humans. It can be greater in machines.

- The language used should have functions and predicates corresponding to its own *abstract syntax*. Lisp syntax will help. So will having concepts and contexts as objects.

- It's not clear whether other improvements in symbolic logic are needed to make self-aware computer systems.

SUMMARY

Here's what AI needs from symbolic logic.

• Systems with concepts as objects and contexts as objects.

• Systems allowing partially defined objects.

• Heavy duty set theory. More generally, systems good for proving theorems within and not morely theorems about. Enough definitions and theorems so that proofs are as short as informal proofs—indeed direct transcriptions of informal proofs.

• Systems whose theories are objects that can be reasoned about in higher level formal theories.

REMARKS AND REFERENCES

- One common reaction to the idea of nonmonotonic reasoning is that it can all be done with probability theory. They're different but related.

Nonmonotonic reasoning is often used to form the propositions to which a Bayesian will ascribe probabilities. An example is the proposition that there are no material objects relevant to a problem except for those whose existence follows from the known facts. In the well known missionaries and cannibals problem this excludes the existence of a bridge or something wrong with the boat.

- The biggest unmet requirement for computer programs to achieve goals using logical AI is the ability to describe domain dependent reasoning strategies, preferably declaratively. This includes resource bounded reasoning.

• Another problem is connecting thinking with lower level computational processes.

• References are be at the URL http://www-formal.stanford.edu/jmc/asl.html.