ADVICE ABOUT LOGICAL AI

John McCarthy
Computer Science Department
Stanford University
Stanford, CA 94305
jmc@cs.stanford.edu
http://www-formal.stanford.edu/jmc/

1 Introduction

I thank the LICS03 program committee for inviting me to talk about nonmonotonic reasoning. I have only two small technical results to report. The rest is history and opinion. I can't give a full history of the developments of nonmonotonic reasoning since the 1970s, because I have read too few of the papers. In some measure this is because the study of nonmonotonic reasoning has become a mathematical subject of its own independent of its applications to AI.

Formalized nonmonotonic reasoning rose in the 1970s in connection with research on artificial intelligence, so I'll express some opinions on what has developed and some advice for the future of nonmonotonic reasoning in AI.

2 Mathematical logic and common sense reasoning

Leibniz, Boole and Frege all had the goal of making a logic that would permit reasoning about common sense phenomena. However, mathematical logic as put in good form by Frege (and Pierce) and polished in the first half of the 20th century, was adequate for formalizing mathematics and other mathematical sciences and inadequate for formalizing common sense knowledge and reasoning. This was because the logicians didn't formalize nonmonotonic reasoning.

The reason is that the logicans' goal was guaranteed correct reasoning, and this is necessarily monotonic. Whenever $A \vdash p$ and $A \subset B$, we also have $B \vdash p$, because the proof of p from A will also serve as a proof from B.

Common sense reasoning is often nonmonotonic. For example, from $Bird(Tweety)$ we wish to infer $Flies(Tweety)$ even though if told about an exceptional bird that doesn't fly because it is a penguin, we don't wish to reach a contradiction. This isn't the place for a substantial exposition of formalized nonmonotonic reasoning, but we will remind the reader of some kinds of applications.

Nonmonotonic reasoning has several uses.

1. As a communication convention. Suppose A tells B about a situation involving a bird. If the bird cannot fly, and this is relevant, then A must say so. Whereas if the bird can fly, there is no requirement to mention the fact. For example, if I hire you to build me a bird cage and you don't put a top on it, I can get out of paying for it even if you tell the judge that I never said my bird could fly. However, if I complain that you wasted money by putting a top on a cage I intended for a penguin, the judge will agree with you that if the bird couldn't fly I should have said so.

The proposed Common Business Communication Language (CBCL) ([2]) must include nonmonotonic conventions about what may be inferred when a message leaves out such items as the method of delivery.

2. As a database or information storage convention. It may be a convention of a particular database that certain predicates have their minimal extension. This generalizes the closed world assumption. When a database makes the closed world assumption for all predicates it is reasonable to imbed this fact in the programs that use the database. However, when only some predicates are to be minimized, we need to say which ones by appropriate sentences of the database, perhaps as a preamble to the collection of ground sentences that usually constitute the main content.

Neither 1 nor 2 requires that most birds can fly. Should it happen that most birds that are subject to the communication or about which information is requested from the data base cannot fly, the convention may lead to inefficiency but not incorrectness.

3. As a rule of conjecture. This use was emphasized in (McCarthy 1980). The circumscriptions may be regarded as expressions of some probabilistic notions such as "most birds can fly" or they may be expressions of standard cases. Thus it is simple to conjecture that there are no relevant present material objects other than those whose presence can be inferred. It is also a simple conjecture that a tool as-

Proceedings of the 18th Annual IEEE Symposium on Logic in Computer Science (LICS'03)
1043-6871/03 $17.00 © 2003 IEEE

serted to be present is usable for its normal function. Such conjectures sometimes conflict, but there is nothing wrong with having incompatible conjectures on hand. Besides the possibility of deciding that one is correct and the other wrong, it is possible to use one for generating possible exceptions to the other.

4. As a representation of a policy. The example is Doyle's "The meeting will be on Wednesday unless another decision is explicitly made". Again probabilities are not involved.

5. As a very streamlined expression of probabilistic information when numerical probabilities, especially conditional probabilities, are unobtainable. Since circumscription doesn't provide numerical probabilities, its probabilistic interpretation involves probabilities that are either infinitesimal, within an infinitesimal of one, or intermediate — without any discrimination among the intermediate values. The circumscriptions give conditional probabilities. Thus we may treat the probability that a bird can't fly as an infinitesimal. However, if the rare event occurs that the bird is a penguin, then the conditional probability that it can fly is infinitesimal, but we may hear of some rare condition that would allow it to fly after all.

Why don't we use finite probabilities combined by the usual laws? That would be fine if we had the numbers, but circumscription is usable when we can't get the numbers or find their use inconvenient. Note that the general probability that a bird can fly may be irrelevant, because we are interested in the facts that influence our opinion about whether a particular bird can fly in a particular situation.

Moreover, the use of probabilities is normally considered to require the definition of a sample space, i.e. the space of all possibilities. Circumscription allows one to conjecture that the cases we know about are all that there are. However, when additional cases are found, the axioms don't have to be changed. Thus there is no fixed space of all possibilities.

Notice also that circumscription does not provide for weighing evidence; it is appropriate when the information permits snap decisions. However, many cases nominally treated in terms of weighing information are in fact cases in which the weights are such that circumscription and other defaults work better.

6. Auto-epistemic reasoning. "If I had an elder brother, I'd know it". This has been studied by R. Moore.

7. Both common sense physics and common sense psychology use nonmonotonic rules. An object will continue in a straight line if nothing interferes with it. A person will eat when hungry unless something prevents it. Such rules are open ended about what might prevent the expected behavior, and this is required, because we are always encountering unexpected phenomena that modify the operation of our rules. Science, as distinct from common sense, tries to work with exceptionless rules. However, this means that common sense reasoning has to decide when a scientific model is applicable, i.e. that there are no important phenomena not taken into account by the theories being used and the model of the particular phenomena.

Seven different uses for nonmonotonic reasoning seem too many, so perhaps they can be consolidated.

2.1 Advice on nonmonotonic reasoning

1. Formalisms requiring fixed point constructions like the full logic of defaults are likely to be difficult to compute with. Default logic restricted to normal defaults seems not to require a fixed point definition.

2. In [5] I do the nonmonotonic reasoning one situation at a time. This works well for projection problems, but is inadequate for the stolen car scenario which postulates a fact about a future situation which is not inferred from facts about previous situations.

3 Logical AI

Here are some observations.

1. *Artificial intelligence.* The long term goal of artificial intelligence research should be human-level AI, i.e. computer programs with at least the intellectual capabilities of humans. There are two main approaches to seeking this goal.

   Biological approaches build agents that imitate features of the physiology or psychology of humans. Most cognitive science agents imitate humans at the psychological level; connectionist systems and their neural net relatives imitate at the physiological level.

   For progress the biological AI researchers need to figure out how to represent facts independent of their purpose, to make systems capable of sequential behavior and to figure out what information to build into their systems corresponding to the rich information that human babies are born possessing.

   Engineering approaches to AI regard the world as presenting certain kinds of problems to an agent trying to survive and achieve goals. It studies directly how to achieve goals. The **logical approach** is a variety of the engineering approach. A logical agent represents what it knows in logical formulas and infers that certain actions or strategies are appropriate to achieve its goals.

   Human level logical agents need at least (1) continued existence over time, (2) improved ability to reason about action and change, (3) more elaboration tolerant formalisms, (4) the ability to represent

Proceedings of the 18th Annual IEEE Symposium on Logic in Computer Science (LICS'03)
1043-6871/03 $17.00 © 2003 IEEE

and reason about approximately defined entities, (5) enough self-awareness and introspection to learn from the successes and failures of their previous reasoning, (6) domain dependent control of theorem provers and problem solvers, and (7) identifying the most basic common-sense knowledge and getting it into the computer. This must include knowledge that people use without being able to formulate it verbally.

The classical AI problems—the frame problem, the qualification problem and the ramification problem [8]—have been solved for particular formalisms for representing information about action and change. However, these currently known representation formalisms are unsatisfactory in various respects. Improved ways of representing action and change may present new versions of the frame, qualification and representation problems.

It's a race to human-level AI, and I think the logical approach is ahead. Why?

The logical approach to AI has faced and partly solved problems that all approaches will face sooner or later. Mainly they concern identifying and representing as data and programs information about how the world works. Humans represent this information internally, but only some of it is verbally accessible.

The logical approach also has the advantage that when we achieve human-level AI we will understand how intelligence works. Some of the evolutionary approaches might achieve an intelligent machine without anyone understanding how it works.

4 Problems of logical AI

What are the problems that must be faced by logical AI before we can reach human level intelligence? Here are a few.

1. *facts about action and change.* This has been the major concentration of work in logical AI. Present formalisms are pretty good for representing facts about discrete sequences of actions and their consequences. Programs exist for the automatic translations of planning problems so described into propositional theories, and the propositional problem solvers are often adequate to solve the problems. The situation is not so good for continuous change, concurrent events, and problems involving the actions of several agents.

2. *elaboration tolerance including the frame problem.* Existing AI systems, including both logical and biological, are extremely specialized in what information they can take into account. Taking into account new information, even information not contradicting previous knowledge, often requires rebuilding the system from scratch. Logical AI has studied elaboration tolerance, a special cases of which is the presented by the frame problem and the qualification problem. The frame problem involves the implicit specification of what doesn't change when an event occurs, and the qualification problem involves elaborating the sufficient conditions for an event to have its normal effect. [8] treats elaboration tolerance and there is also my http://www-formal.stanford.edu/jmc/elaboration.html. [1] implements 9 of my 19 elaborations in the McCain-Turner causal calculator. J. Gustafsson and J. Kvarnström got 14 of the 19.

3. *nonmonotonic reasoning.* Humans and machines usually reach conclusions on the basis of incomplete knowledge. These conclusions may change when new knowledge becomes available. Probability theory is applicable to a part of the problem, but more general nonmonotonic logics seem to be also required.

4. *the three dimensional world: approximate knowledge.* The knowledge available to a person or robot of its three dimensional surroundings and of the objects it needs to manipulate is almost always very approximate, though sometimes precise geometric models like rectangular parallelepiped approximate real objects well enough. Logical AI needs a general theory of approximate objects.

5. *the relation between appearance and reality.* We and any robots we may build live in a world of three-dimensional objects built up from substances that are in turn made of molecules. Our direct information about this world comes directly from senses like vision and hearing that only carry partial information about the objects. We, even babies, are built to infer information about the three dimensional objects from observations and from general information about what kinds of objects there are.

   Machine learning and most AI have treated classifying appearances but don't go beyond appearance to the reality. I have a puzzle about this at http://www-formal.stanford.edu/jmc/appearance.html.

6. *the grounding of logic in partially understood phenomena*. I don't yet have a precise formulation of this new (to me) idea, so I'll proceed with an example. Suppose I wish to get my car keys or my pocket knife from my pants pocket, or we wish to program a robot to get objects from pockets. I put my hand in my pocket and fumble around until one of my fingers encounters one of the surfaces of the desired object. Moving my fingers to grasp the object is another fumble, because the

Proceedings of the 18th Annual IEEE Symposium on Logic in Computer Science (LICS'03)
1043-6871/03 $17.00 © 2003 IEEE

fingers may encounter other objects or the cloth of the pocket.

It is not reasonable to attempt to fully axiomatize the process of fumbling in the pocket in the sense of giving the effect of each motion of the hand and fingers. The axioms need only go down to the point of assuring the person or robot that a recognizable surface of the desired object will be encountered. The effects of the program that fumbles in the pocket may be axiomatized, but the fact that the program works is not derived from axioms about pockets and their contents but is an empirical fact. Perhaps more can usefully be expressed axiomatically about the motion of the hand in the pocket, but I don't know how to write these axioms.

Many of the axioms about the effects of actions are known empirically rather than derived from facts about the effects of lower level actions and other events.

7. *free will*. What a robot should believe about its own abilities, i.e. about the choices available to it was first discussed in [7]. Here is a formula for free will in a deterministic universe based on the situation calculus formalism of [6].

$$
\begin{array}{l}
Occurs(Does(John, \\
\quad \textbf{if} \\
\text{Thinks-better}(John, Result(Does(John, a1), s), \\
Result(Does(John, a2), s)) \\
\quad \textbf{then } a1 \\
\quad \textbf{else } a2 \\
\qquad), s).
\end{array}
$$

Here $Occurs$ tells what happens and is thus the deterministic part. $Result(a1, s)$ and $Result(a2, s)$ are the two possible actions and are thus the free will part. Thinks-better gives the decision criterion. Overall the formula says that John does the action that has result he likes best.

Logical AI has faced the above phenomena. Other approaches haven't faced them explicitly. Maybe it will turn out that they don't have to face them, but I have seen no arguments to that effect.

My views on many of the specific problems are in articles published in various journals; almost all are to be found on my web page
`http://www-formal.stanford.edu/jmc/`.
There is no complete summary, but
http://www-formal.stanford.edu/jmc/whatisai.html and [3] express the logical AI point of view. [8] covers much of the logical AI approach.

No approach is close to human-level AI and or within development range. No-one can convincingly say that given a billion dollars he could reach human-level. A critical level of AI will be reached, as Douglas Lenat pointed out, when an AI system has enough basic common sense information to be able to get more information by reading books.

Human-level intelligence is a difficult scientific problem and probably needs some new ideas. These are more likely to be invented by a person of genius than as part of a Government or industry project. Of course, present ideas and techniques are sufficient for many useful applications.

This research was partly supported by ARPA (ONR) grant N00014-94-1-0775.

References

[1] V. Lifschitz. Missionaries and cannibals in the causal calculator. In A. G. Cohn, F. Giunchiglia, and B. Selman, editors, *KR2000: Principles of Knowledge Representation and Reasoning,Proceedings of the Seventh International conference*, pages 85–96. Morgan-Kaufman, 2000.

[2] J. McCarthy. Common Business Communication Language[1]. In A. Endres and J. Reetz, editors, *Textverarbeitung und Bürosysteme*. R. Oldenbourg Verlag, Munich and Vienna, 1982.

[3] J. McCarthy. Artificial Intelligence, Logic and Formalizing Common Sense[2]. In R. Thomason, editor, *Philosophical Logic and Artificial Intelligence*. Klüver Academic, 1989.

[4] J. McCarthy. *Formalizing Common Sense: Papers by John McCarthy*. Ablex Publishing Corporation, 1990.

[5] J. McCarthy. Actions and other events in situation calculus. In B. S. A.G. Cohn, F. Giunchiglia, editor, *Principles of knowledge representation and reasoning: Proceedings of the eighth international conference (KR2002)*. Morgan-Kaufmann, 2002. http://www-formal.stanford.edu/jmc/sitcalc.html.

[6] J. McCarthy. Deterministic free will. http://www-formal.stanford.edu/jmc/freewill2.html, October 2002.

[7] J. McCarthy and P. J. Hayes. Some Philosophical Problems from the Standpoint of Artificial Intelligence[3]. In B. Meltzer and D. Michie, editors, *Machine Intelligence 4*, pages 463–502. Edinburgh University Press, 1969. Reprinted in [4].

[8] M. Shanahan. *Solving the Frame Problem, a mathematical investigation of the common sense law of inertia*. M.I.T. Press, 1997.

[1] http://www-formal.stanford.edu/jmc/cbcl.html
[2] http://www-formal.stanford.edu/jmc/ailogic.html
[3] http://www-formal.stanford.edu/jmc/mcchay69.html

Proceedings of the 18th Annual IEEE Symposium on Logic in Computer Science (LICS'03)
1043-6871/03 $17.00 © 2003 IEEE