REASONING ABOUT KNOWLEDGE IN ARTIFICIAL INTELLIGENCE

Robert C. Moore
SRI International
Menlo Park, California, and Cambridge, England

SUMMARY

This talk looks at work on reasoning about knowledge within what might be termed "classical" AI. That is, it is assumed that the information contained in an intelligent system is for the most part embodied in data structures that explicitly represent the propositions that the system knows or believes, and that queries to this "knowledge base" are handled by applying inference rules in a way that amounts to searching for a derivation of an answer to the query.

Given this point of view, the first question that comes up in reasoning about knowledge in AI is how to *represent* information about what someone knows. The fact that this is not altogether obvious has been a prime motivation for work in AI on reasoning about knowledge. Another motivation comes from AI work on planning. In trying to formulate a plan of action to achieve some goal, an agent may not have enough information. It is often necessary to reason about what knowledge is needed to carry out a plan and how that knowledge can be obtained. A third motivation is found in recent work in natural-language processing that tries to take into account the mental state of the person that the system is communicating with in interpreting and generating utterances. This usually requires reasoning about what that person knows. Finally, there are connections to work on nonmonotonic reasoning -- most nonmonotonic reasoning systems have special inference rules that do not apply if their conclusions are known to be false, and in some systems this involves explicit reasoning about what the system itself knows.

A number of different techniques have been used or proposed for reasoning about knowledge in AI systems. Several different attempts have been made to face the problems head on, reasoning explicitly in a first-order way, treating propositions as objects and knowing as a binary relation. Another approach that has been extensively explored is to axiomatize within first-order logic the possible-world semantics of some variant of Hintikka's modal epistemic logic. A third approach uses the idea of *semantic attachment* to simulate other agents' reasoning from their knowledge.

Two particular issues deserve special comment. One is the question of why the application of the known decision procedures for various modal epistemic logics is not among the list of commonly used techinques for reasoning about knowledge in AI. The principal answer is that decision procedures generally exist only for *propositional* modal epistemic logics, and many AI applications seem to require quantification. Secondly, the straightforward application of these decision procedures in many cases would require treating the system's entire knowledge base as the antecedent of the formula to be decided. Practical application of such decision procedures would thus require that only the *relevant* information in the knowledge base be taken into account, and determining relevance is a notoriously unsolved problem.

Perhaps the most hotly contested issue in this field, however, is the issue of logical omniscience. Most approaches based on modal epistemic logic or possible worlds have built into them the assumption that agents know all the logical consequences of their knowledge. This assumption, though enormously simplifying, is generally regarded to be clearly false. Nevertheless, it is an assumption that is frequently made for lack of an alternative that permits a system to conclude that other agents will make all the inferences that seem to be obvious.