WHAT IS COMMON SENSE

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

2002 Feb 14, 3:21 p.m.

Abstract

1 Introduction

There is common sense knowledge and common sense ability. Common sense knowledge includes facts about events occurring in time, about the effects of actions by the knower and others, about physical objects and how they are perceived, and about their properties and their relations to one another. An example is the fact that eggs contain a yolk and a white and a shell, how to recognize an egg, the effects of hard boiling them and the effects of dropping them. Common sense ability involves the use of common sense knowledge and the observation of the world to decide what to do to achieve one's goals. The "common" in "common sense" refers to the fact that a large amount of this knowledge and ability is common to all humans. Not much of it is understood well enough to include it in computer programs.

For AI purposes we take common sense to include certain knowledge about the world and a certain ability to use it to infer addtional facts from observations and general knowledge and to use this reasoning to achieve goals.

The core of common sense knowledge concerns actions and other events. Every human knows that he can perform a variety of actions and that the chosen actions affect the future in characteristic ways. However, if we want computer programs to know this, we must provide the knowledge and build into their structure the ability to use it. Very few expert systems are provided with any such knowledge. Instead they have rules that say what action to take when given conditions occur. The likely consequences of these actions were considered by the programmer in deciding what rules to provide, but the languages for writing expert systems provide no direct way of his telling the program such facts or any automatic way for the program to use the information if he did.

Humans have both. If the car in front of us suddenly stops, a learned rule causes us to automatically hit the brakes. We also have at least a verbal knowledge of what will happen if we don't. Some of us can think fast, so that we can quickly decide whether it is better in a particular case to hit the brakes as hard as possible or try to swerve around the car in front. This requires estimating the chances of a head on collision with a car going the other way.

It is not just an oversight that programs don't have both. Rules are easier to write than common sense facts, and it is very much easier to make a system that obeys rules or compiles them into program than to make a system that acts according to reasoning from facts or compiles facts into rules or into program. It would be still harder to make a system that will compile rules that take facts into account but also provide for reverting to reasoning when this is appropriate, i.e. when the situation is complex and there is time to think. Humans have, and machines need, the ability to operate by rule and intermittently monitor and modify the operation of the rules.

The facts that describe the consequences of actions, and while we're at it, the facts that describe the consequences of events the actor doesn't control, are the most important common sense knowledge. However, there's lots more.

The core of the knowledge concerns events that occur in time and the effects of events on the state of the world. Knowledge about the effects of actions by the knower are particularly important. Facts about the relations between appearances and reality are included.

Facts about goals and how to achieve them.

Facts about what actions are permissible.

To what extent are common sense abilities achievable by production sys-

tems?

We would like a system that can find an algorithm for block stacking given the facts about moving blocks. The ontology of this system should include sequences of events. This enables us to say that an efficient algorithm is one whose execution results in a short (more generally cheap) sequence of actions and other events leading to the goal.

What is common sense?

It is now generally agreed among AI researchers that making programs with common sense abilities is at present the key problem facing AI research. However, there is so far no paper attempting a general description of what common sense ability is. There are many examples of specific common sense capabilities. The object of this paper is to attempt a general description of common sense.

Common sense involves certain abilities to decide what to do to achieve goals. These abilities are common to all non-feeble-minded humans; indeed many very retarded people possess common sense abilities so far lacking in any computer program. Many of these abilities seem to transcend specific domains of application.

Human activity seems to be governed by a combination of behavior rules and reasoning based on facts. By a behavior rule we mean one of the form: If X is true, then do Y. X is typically a conjunction and included in the conjunction are usually statements about current goals. Moreover, both X and Y contain variables, and the variables have to be matched against the world and the database in order to generate a definite action.

However, besides rules, humans also accumulate facts. Here are some examples:

1. London is in the south of England.

2. When objects collide they usually make a noise.

3. The Stanford computer science comprehensive examination contains a section on numerical analysis.

4. When a person fires a handgun a chemical residue is left on his hands.

Such facts may be used to answer geography questions, to conclude that a side trip to London while visiting Edinburgh may take more time than is available, to conclude that a screech of brakes was not actually followed by the collision of the cars, to decide to take one's graduate work in computer science elsewhere than in Stanford or to decide to have someone else shoot one's desired victim. The writer of a rule-based expert system will take such facts into account in writing his rules. However, the present expert system

technology has only a limited ability to include the facts themselves and insure that they will be taken into account in as wide a variety of situations as a human would do.

We may call the human ability to take facts into account common sense reasoning ability. In any given case, it may be difficult to ascertain whether an action rule is being invoked or whether reasoning with facts is being done. The distinction is not always clear. First a system that reasons with facts may itself be built up from rules. Second the information is available in redundant forms.

Consider a person whose car needs fixing. He may decide that he will be without his car for several days and hence should shop for several days food in advance, because it will inconvenient to do so without the car. It seems to me that for some people this is an example of fact based reasoning, but it may become "compiled" into a rule after some experience. Some smart people will compile it into a rule from just contemplating the possible situation before any experience.

Rule based systems are usually considered "brittle". By comparison with humans, especially the experts whose abilities they are designed to incorporate, they cannot take new facts into account unless they fit the format of their databases. For example, it might be difficult to tell Mycin about AIDS without rewriting it extensively. Presumably, if AIDS is a possible diagnosis, the probabilities of the different bacterial infections are drastically altered, because some bacteria that the human body can ordinarily handle without help are now dangerous.

Often facts enter generally rule-based behavior as an exception handling mechanism. One is carrying out some routine activity in pursuit of a customary goal when suddenly an event not previously imagined occurs and one says to oneself, "I'd better think". Thinking often involves going into a fact-dominated mode. Thinking is often unsuccessful in the sense that one later concludes that one could have handled a situation much better if only one had thought of the answer in time. A good robot will presumably have many of the same characteristic. It has rules, but when exceptions are encountered, it has to reason, and this reason is only sometimes capable of coming up with the right answer even when the relevant facts are all in its database.

The rest of this essay is devoted to two questions. First, what are the main common sense facts? Second, how are these facts used in common sense reasoning in conjunction with rules. Only very preliminary answers are possible in the present state of AI.

THE FACTS OF THE COMMON SENSE WORLD

It seems that the most important class of facts are those that permit the prediction of the effects of events including actions. Basic to this is the fact that the world does change and that actions and other events have effects. However, this fact is so basic that its explicit representation may not play an important role. Namely, it is built into the human mechanism rather than used as a fact by a reasoning mechanism. Such treatment may also be appropriate in designing computer programs. An analogous example may be helpful.

Consider the regress involving modus ponens proposed by Lewis Carroll and which Douglas Hofstadter has made much of. Achilles infers q from p and $p \rightarrow q$. The tortoise asks why, and Achilles cites the rule, "If p is true and $p \rightarrow q$ is true, then q is also true". The tortoise asks why, and the regression continues. Achilles should have stopped the regress at the first stage. When asked why he thinks so, he should have answered, "Because that's the way I'm constructed. I'm built to allow modus ponens". The higher level modus ponens's aren't higher level principles, they are simply more examples of the same principle.

In any case, the idea that the future lies ahead and can be affected is most likely built in. The fact that we can formulate it as a fact is of secondary importance in daily life, although it is important when thinking about philosophy or about designing robots.

There is an important difference between common sense ability to predict the consequences of physical actions and the laws of physics. Namely, common sense physics allows only incomplete partial prediction. When an object is thrown, we can predict where it will land very approximately, but this information is insufficient to predict where it will go next.

A leftover: The system that accumulates facts and takes them into account may itself be organized as a rule based system. Moreover, a system organized primarily as rule a based system, i.e. doing do most of its work by executing rules, may go into a mode that uses facts when exceptions are detected. A person may say, "This has gotten complicated. Now I have to think."