Long abstract:

The role of temporal reasoning subsystems in the architecture of autonomous robots

Erik Sandewall

Department of Computer and Information Science, Linköping University, 58183 Linköping, Sweden.
E-mail ejs@ida.liu.se.

TEMPORAL REASONING SYSTEMS

A *temporal reasoning system* is a software or an algorithm which performs reasoning tasks in a temporal logic. Several such reasoning tasks need to be performed in autonomous robots, in particular *planning*, *postdiction* for the purpose of diagnosis, *replanning* to deal with unexpected situations, and *prediction* which gives the robot a capability to anticipate the future and thereby to react in time for avoiding dangers. Today these reasoning tasks are often conducted by ad hoc methods: specialized planning algorithms, simulation programs for prediction, and so on. However this state of affairs is not satisfactory in the long run. Ideally one should have a common basis in the form of an adequate temporal logic, so that the various temporal reasoning tasks can be defined in terms of the same logical base.

The emergence of a common logical basis does not necessarily mean that a single implementation will be practicable, for example through a uniform theorem prover. For performance reasons, and in particular in order to satisfy real-time requirements, the logical basis will more likely only be used for the purpose of specification. Several distinct and specialized software subsystems can be developed for planning, prediction, and other functions within the robot, but on the same theoretical basis.

The research problem of developing temporal reasoning subsystems for autonomous robots in a systematic fashion is therefore decomposed into several parts: the identification of an adequate temporal logic, the definition of the various inference operations (planning, postdiction, etc.) in terms of the logic, and the identification of adequate algorithms and reasoning methods for each inference operation.

In a monograph[?] that is about to come out as a book, I present a new approach to these problems, with an emphasis on the first one. Any choice of temporal logic for this purpose must deal with the classical "frame problem" in A.I. Rather than proposing one more solution to the frame problem, I define and use a framework for assessing the range of applicability of such logics. The key idea is to first define an *underlying semantics* which defines what are the intended models for a given scenario description. This underlying semantics is used for defining a *taxonomy* of reasoning problems, based on what ontological and epistemological assumptions are being made. For a number of proposed logics, it is then possible to identify a class of reasoning problems such that the set of models which are selected or preferred by the logic equals the set of intended models according to the underlying semantics. The taxonomy of reasoning problems is then used as the "coordinate system" whereby the range of applicability can be precisely defined and formally proved.

From this analysis we have determined that the range of correct applicability for several "standard" methods is surprisingly narrow. The identification of the restrictions has however made it possible to find other methods, with modified non-monotonic entailment conditions, whose range of applicability is considerably broader. At the same time one should notice that a broad range of applicability is likely to have a price in terms of performance. For a given application domain, it is reasonable to first identify what class of reasoning problems are required by the application, and then to select that entailment method, among the adequate ones for the application at hand, which has the most efficient implementation to go with it.

After the precise analysis has been obtained for the entailment methods, and thereby for the non-monotonic logics for the "frame problem", the next step is to define the inference operations in precise terms. For some of them such as prediction and postdiction this is more or less trivial, but for others such as the planning operation the matter is a bit more complex. We are beginning to obtain results also for these questions in the research of our group.

AUTONOMOUS AGENT MODELS IN THE UNDERLYING SEMANTICS

What has been described so far suggests that the logic is the primary construct, and that it is used for building some of the modules in the actually implemented robot system. However there is also a reverse relationship between the logic and the robot architecture, because the underlying semantics is defined in terms of a formally defined autonomous agent i.e. an idealized robot.

The basic idea is as follows. The robot is assumed to have a conventional, three-layer architecture, where the uppermost layer is in charge of reasoning, and corresponds to Newell's *knowledge level*. Let us use the term *ego* for the uppermost level, and *world* for the combination of the robot's environment and its own two lower levels.

The interface between the ego and the world is then one of the significant design decisions in the design of the actual robot. However that interface also turns out to have a formal significance. In the underlying semantics, the set of intended models for a given scenario description is defined in terms of a *game* between the ego and the world. The game starts when time equals zero, and ends when time has run over the integers to plus infinity. The ego and the world make alternating moves. When the ego has the move, it may invoke actions (i.e. require the lower levels of the robot architecture to perform the action), but time does not advance as a result. When the world has the move, it actually performs the actions that have been invoked by the ego, and advances time accordingly.

For a given scenario description, the set of intended models is defined as follows. One first identifies the world (as a game player) which is exactly defined by the scenario description. Then one obtains the set of all possible games, for arbitrary initial state and arbitrary ego behavior. Finally one restricts the set of games to those that satisfy the given scenario description, modulo also the given epistemological assumptions with respect to complete knowledge about some aspects of the scenario.

More specifically, we use a *trajectory semantics* where the moves of the world-player are defined as follows. Actions are assumed to occur sequentially. For each combination of an action and a current state of the world, there is a set of trajectories which specify possible ways of performing the action. Each trajectory is a finite sequence of partial states of the world, specifying the successive values of those fluents which may change as a result of the action. Other fluents are assumed to be inert – this is how the assumption of inertia is built into the underlying semantics.

When the world-player is to move, it therefore identifies the current state of the world, r, and the action A that has currently been invoked by the ego. It then selects one arbitrary member of the set of trajectories for the given r and A, and extends the present finite history using that trajectory.

What has here been described in brief outline is the formal machinery whereby we can define the taxonomy of reasoning problems, define the set of intended models for a given scenario description, and analyze the range of applicability for a number of proposed non-monotonic logics.

The trajectory semantics characterizes the ego-world relationship on a discrete level. A necessary next step of research will be to show how the trajectory semantics can be reconstructed from a more precise, often quantitative "model" of the robot's environment combined with a "model" of how the robot perceives that environment.

MUTUAL RELATIONSHIP BETWEEN AGENT ARCHITECTURE AND LOGIC

In summary, I have tried to show that there is a two-way relationship between the temporal logic and the architecture of the robot or autonomous agent. On a concrete level, the actually implemented robot contains subsystems for temporal reasoning which have been specified in terms of an temporal logic. However on a more abstract level, the robot architecture is also a formal device which is used as the basis for the definition and assessment of the proposed logics. I believe that in this way a rigorous and coherent basis for continued theoretical research about autonomous robots will evolve.