In NASA WRAC 2001.

Evolution of Agent Architectures

Henry Hexmoor

Computer Science & Computer Engineering Department
Engineering Hall, Room 313, Fayetteville, AR 72701
{hexmoor}@uark.edu

Agent architectures provide the blueprints for the design and development of individual agents. The purpose of agent architecture is to define modes and modules of agent's interaction in the world as well as connections among internal components of an agent. An agent can be understood in terms of its architectural description of its perceptual, deliberation, and acting capabilities. Architectures have been used to describe robot control software [Musliner, Durfee, and Shin, 1993]. Such architectures emphasized rapid processing of early perception and low-level actuation capabilities. Brooks' subsumption architecture has been an influential guide for fast and reactive robotic actions [Brook 1986]. Although subsumption was good for real-time processing needed in robotics, it never became useful for agents. This is partly because being reactive is a standard property of agents. Pro-active architectures such as logic-based, utility-based, or belief-desire-intention (BDI) have been more popular in agent architectures [Wooldridge, 2000]. Intentional agents are modeled in multi-modal BDI logics. Each architecture has its strengths and weaknesses that make it suitable for particular roles or particular types of problems. Instead of comparing and contrasting architectures, here we give a partial list of collective properties for pro-active architectures: logical correctness, ability to initiate a new course of action, ability to form and manipulate explicit intentions, ability to reason based on nontrivial models of the world, ability to adapt, ability to learn, ability to interact with emotions, and ability to react when there is not enough time to complete reasoning. In nontrivial monolithic systems, proactive architectures addressed many issues including: world modeling, modularity of cognitive functions (such as planning and learning), affect, and uncertainty management. For agent-based systems, reasoning about autonomy is a specific area of concern. Finally, nontrivial agents who have to account for other agents and be social must address many issues. A partial list is: coordination, cooperation, teamwork, and other relationships and social attitudes such as autonomy, veracity, awareness, benevolence, rationality, roles, obligations.

More complex agents must exhibit self-awareness, self-modifying behavior, dealing with time-constraints, and nontrivial social actions. Once an agent faces time and resource limits, it becomes important for the agent to incorporate into its architecture the ability to model its own capabilities and limitations, along with modeling the external environment. Principled adaptation techniques are needed to improve performance based on experience. These adaptation methods are also needed to improve self-modifications to satisfy human

users. Verification and validation of performance within guaranteed time-critical boundaries are needed in mission critical applications. Agents operating under uncertain conditions need to formulate policies to do the best they can in response to the uncertain unfolding of their environment. Resource-limited and time-constrained agents need to balance responsiveness with farsightedness. Functionally independent plug and play design of agent architectures will be useful to install or remove components that allow favorable behaviors that are predictable in large time scales. Self-modifying agents need to make changes that are justifiable and still ensure sufficient predictability in behaviors upon which operators depend. Rapid reasoning is required about resource allocation to competing demands as well as preserving missions and goals in which they are deployed.

Agents are being deployed in ever more common as well as complex places to take over dirty, dull, and dangerous tasks. This requires agents to be aware of the operating context of their environment and to work sociably and responsibly with humans in the environment. Agents need to automatically acquire user preferences to form valid probabilistic projections about future courses of events and the desirability of each of them to the human whose interests are being represented. Much progress has been made along many of these fronts, but much remains yet to be done. In particular, further work should be strongly encouraged in developing agents that are capable within the constraints of dynamic and uncertain application domains.

Acknowledgements

This work is supported by AFOSR grant F49620-00-1-0302.

References

1. R. Brooks, 1986. *A Robust Layered Control Systems for a Mobile Robot*. IEEE Journal of Robotics and Automation, RA 2-1:14-23.
2. D. Musliner, E. Durfee, and K. Shin, 1993. *CIRCA: A Cooperative Intelligent Real-Time Control Architecture*. IEEE Transactions on Systems, Man, and Cybernetics.
3. M. Wooldridge, 2000. *Reasoning about Rational Agents*, The MIT Press.