CPS331 Lecture: Agents and Robots   last revised November 18, 2016

Objectives:

1. To introduce the basic notion of an agent
2. To discuss various types of agents
3. To introduce the subsumption architecture for robots

Materials:

1. Projectable of diagram showing PAGE
2. Projectable of Russell and Norvig figure 2.3
3. Projectable of Russell and Norvig figure 2.13
4. Projectables of Russell and Norvig figures 2.7, 2.9, 2.11, 2.12
5. Projectable of Blackboard architecture example (Rich/Knight figure 16.2)
6. Projectables of centralized and subsumption robot architectures - Figures 1 and 2 from Brooks paper

I. **Introduction**

   A. Definition: An agent is a computational system that performs some task with a measure of autonomy. An agent has appropriate means of perceiving its environment (sensors) and acting upon its environment (effectors).

      1. Autonomy is a key characteristic - what distinguishes an "agent" from a more conventional program such as a word processor or web-browser.

         a) An agent is typically given a task to perform on behalf of a user - either a human user, or another agent that uses its services to accomplish its goals. It does so with minimal interaction with that user - though it may eventually report its discoveries back to its user.

            Examples: web search agents, GPS

         b) As noted, in many cases agents may need to interact with other agents to perform their task.

2. An agent typically functions in a complex and changing environment, and is expected to adjust its behavior appropriately.

3. Russell and Norvig suggest that an agent can be described using the acronym PAGE (Percepts, Actions, Goals, Environment)

   PROJECT Diagram of PAGE

   PROJECT Russell and Norvig figure 2.

   A key concept is the notion of a percept sequence - the sequence of information an agent receives from its sensors

B. Two broad types of agents

1. Pure software agents - "softbots"

2. Hardware agents - "robots"

C. Some examples:

ASK

1. Softbots:

   a) spam filters and email distribution agents that route incoming email to appropriate boxes and/or notify the user of email from an "interesting" sender

   b) detection of viruses and other kinds of attack. (Actually, for that matter a virus or worm is also a sort of softbot!)

   c) algorithmic stock trading software

2. Robots:

   a) NASA has used robotic rovers on the moon and Mars. While humans are still involved in making decisions about what to

explore, movement needs to be autonomous due to long signal propagation times.

    b) Police robots: Google search on "police robot"

D. Actually "agent" is an overarching term - an agent is a system that combines appropriate AI methods to fulfill its particular role.

    1. The notion of an agent has emerged in AI as an alternative to classic models of strong intelligence.

    One concept that underlies this is a notion first espoused in a 1986 book by Marvin Minsky called *The Society of Mind* - which he states this way: "we'll show you that you can build a mind from many little parts, each mindless in itself" (quoted on Lucci p. 199)

    2. Some AI books have used the notion of an "agent" as a unifying theme

    Example: the best-selling AI text (Russell and Norvig: *Artificial Intelligence - a Modern Approach*) includes the following phrase in its description: "It uses an intelligent agent as the unifying theme throughout"

    3. We consider agents at this point as a sort of summary of weak AI.

    (We will focus on software aspects, not the hardware issues that arise with robots - though we will talk a bit about one issue that arises in conjunction with robotics).

        a) Agents may make use of any of the techniques we have discussed in this course - e.g. search, planning, use of natural language etc.

        b) Often agents are learning agents - that is, they don't simply perform a pre-programmed task, but learn to perform a task.

II. **Properties of Environments**

An agent is designed to function in a particular type of environment. These environments can be characterized in a various ways. (we'll illustrate using different types of game, but the notion is general)

A. Accessible vs inaccessible - do the agent's sensors provide access to the <u>complete</u> state of the environment (that is everything that is relevant to the agent's decision-making)?

    1. Accessible: most board games

    2. Inaccessible: poker

B. Deterministic vs nondeterministic: is everything that happens in the environment a result of the actions of the agents (note plural)?

Note: an inaccessible environment may appear nondeterministic even though it isn't!

    1. Deterministic: most board games (assuming we regard all the players as agents)

    2. Non-deterministic:

        a) backgammon (use of dice)

        b) poker (inaccessible therefore appears nondeterministic - and also depends on the randomness of the shuffling of the deck.)

C. Episodic vs nonepisodic: does time divide into episodes involving the agent perceiving, deciding and acting, with the agent's decision based only on what it perceives in that episode, or do episodes depend on agents' actions in previous episodes - or is time is continuous? (Note: if a task is non-episodic then the agent doesn't have to "plan ahead")

    1. Episodic: games are usually not episodic since they involve planning strategy. Tasks like part-picking on an assembly line are episodic.

    2. Non-episodic: board games with any strategy aspect; ping-pong - continuous

D. Static vs dynamic: Can the environment change while the agent is deciding? (If the environment is static, the agent can sense the environment and then decide what to do, but if it is dynamic the agent needs to keep sensing while deciding

    1. Static: most board game

    2. Dynamic: capture the flag

E. Discrete vs continuous: is there a discrete set of possible percepts and actions?

    1. Discrete: most board games

    2. Continuous: most games of skill

F. Russell and Norvig illustrate these for a wide variety of types of problem

PROJECT Russell and Norvig Figure 2.13

III. Basic Types of Agents

A. Table-driven agents

    1. We could imagine a table-driven agent being created to function in a static, episodic environment. Each decision the agent makes would be made by looking up its current percepts in a table of specifying the correct action to take for that percept.

    2. Example: a single individual in the population of a genetic algorithm may encode a potential solution to the problem in the form of a table of possible responses to various percepts - e.g. the genetic NIM learner in which each individual was represented as a table of the number of pieces to take for each possible pile size.

3. Of course, such agents are rarely practical since they would require a table that covers every possible percept sequence. (E.g. you could imagine creating such an agent to play tic-tac-toe, where the agent senses the complete current state of the board - but a table-driven chess-playing agent would likely be unfeasible and a table-driven ping-pong playing agent would be impossible.

B. Reflex agents

1. The agent bases its decision completely on its current percepts, but chooses an action by following rules or by some sort of computation

   PROJECT: Russell & Norvig figure 2.7

2. Example: a neural classifier that has been trained functions as a reflex agent when actually being used.

C. Agents with internal state

1. The agent maintains some internal state and bases its decisions on both its state and its current percepts. Thus, the agents are shaped by both its history and the current state of the world.

   PROJECT: Russell & Norvig figure 2.9

2. Example: the various "isa" programs we have worked with responded to user queries based both on the query and an internal state that recorded previous assertions

D. Agents with goals

1. In addition to having state, the agent has one or more goals that it is seeking to achieve, and bases its decisions on its goals, its state, and its current percepts.

   PROJECT: Russell & Norvig figure 2.11

2. Example: the various planners we have worked with

E. Utility-based agents

1. The agent has multiple, possibly conflicting goals, and has some means of comparing the relative "goodness" of different states in order to handle tradeoffs between goals.

   (The technical term for "goodness" is utility.)

   PROJECT: Russell & Norvig figure 2.12

2. Example: a chess-playing program might actually work on several different goals:

   a) Checkmate the opposing king

   b) Keep its own king out of check

   c) Capture an opponent's piece where possible

   d) Keep its own pieces from being captured.

      The static evaluation function is the means by which it compares the utility of various board states in terms of these goals.

3. Example: the ordering of pages presented in response to a query by a search engine like Google

   a) Relevance to the query

   b) Google uses a "page rank" algorithm to give preference to pages that are deemed more reliable or useful on the basis of references to them from other pages.

   c) The first results returned by a query often return references to a few pages from paying advertisers.

F. Additional Types of Agents (not mutually-exclusive with the above)

1. Mobile agents:

   a) General term for an agent that can move from place to place in accomplishing its goals

   b) May be mechanical mobility (e.g. a mobile robot)

   c) May be software mobility (e.g. an agent program that travels from computer to computer)

   Examples:

   ASK

   (1) Web-crawlers used by search engines

   (2) Various kinds of purchasing assistants that access and pool information from other sites (e.g. travel systems like Orbitz)

2. Interface agents: an agent that serves as a sort of personal assistant, carrying out assigned tasks (with some measure of autonomy) on behalf of a user

   Examples:

   ASK

   a) Email distribution systems

   b) Calendar systems

3. Multiagent systems; collaborative systems: Systems in which multiple agents work together to accomplish a common goal (that perhaps no one agent is capable of).

   [The "swarm intelligence" idea fits here]

IV.Agent Architecture

A. An important question in designing an agent is the overall architecture of the agent, and also of the way the agents work together.

B. One early agent-style architecture providing a mechanism for the agents to work together is the blackboard architecture, which was first used in the HEARSAY-II speech recognition system developed in the 1980's.

   1. This system consisted of a number of independent "knowledge sources that considered the original input data and/or hypotheses advanced by other knowledge sources.

   2. These knowledge sources used a shared blackboard as a means of communication. Each agent would "write" its hypotheses on the blackboard for consideration by other agents.

      PROJECT: Blackboard example

   3. The overall goal was to arrive at a single high-level hypothesis that explained the utterance (i.e. what was said)

C. This question is particularly important in conjunction with work in robotics.

   1. Historically, robots were designed using a CENTRALIZED ARCHITECTURE.

      PROJECT - Figure 1 from [Brooks 1985]

      a) Such an architecture is based on a functional decomposition of the overall task - the overall task is broken into models based on the various functions needed (perception, modeling, planning ...)

b) This implies that (at least a portion of) each slice must be implemented before anything can be run

c) This implies that improving the functionality of any piece may result in changes that have to be propagated through the remaining pieces

2. Many robotic systems are based instead on a an architecture like the SUBSUMPTION or BEHAVIOR-BASED architecture first proposed by Rodney Brooks

   PROJECT - Figure 2 from [Brooks 1985]

   a) Such an architecture is based on a decomposition of the overall task in terms of levels of competence.

   b) Each levels subsumes the behavior of lower levels.

   c) New levels of competence can be added on top of existing levels without altering them.