Intelligent Agent

Introduction
Objectives

• The course is aimed at students who already have introductory knowledge of AI.
• Uses multiagent systems as a means to introduce students to some key advanced AI topics.
• Equips students with knowledge to understand and build practical multiagent systems.
Readings

• Textbook:

• Reference books:
Topics

- Multiagent systems
- Intelligent Agents
- Deductive reasoning agents
- Practical reasoning agents
- Reactive and hybrid agents
- Multiagent interactions
- Reaching agreements
- Communication
Assessment

- Class participation: 10%
- Assignments: 30%
- Final Exam: 60%
Overview

• Five ongoing trends have marked the history of computing:
  – *ubiquity*;
  – *interconnection*;
  – *intelligence*;
  – *delegation*; and
  – *human-orientation*
Ubiquity

• The continual reduction in cost of computing capability has made it possible to introduce processing power into places and devices that would have once been uneconomic.

• As processing capability spreads, sophistication (and intelligence of a sort) becomes ubiquitous.

• What could benefit from having a processor embedded in it…?
Interconnection

• Computer systems today no longer stand alone, but are networked into large distributed systems.
• The internet is an obvious example, but networking is spreading its ever-growing tentacles…
• Since distributed and concurrent systems have become the norm, some researchers are putting forward theoretical models that portray computing as primarily a process of interaction.
Intelligence

• The complexity of tasks that we are capable of automating and delegating to computers has grown steadily

• If you don’t feel comfortable with this definition of “intelligence”, it’s probably because you are a human
Delegation

• Computers are doing more for us – without our intervention
• We are *giving control* to computers, even in safety critical tasks
• One example: fly-by-wire aircraft, where the machine’s judgment may be trusted more than an experienced pilot
• Next on the agenda: fly-by-wire cars, intelligent braking systems, cruise control that maintains distance from car in front…
Human Orientation

• The movement away from machine-oriented views of programming toward concepts and metaphors that more closely reflect the way we ourselves understand the world
• Programmers (and users!) relate to the machine differently
• Programmers conceptualize and implement software in terms of higher-level – more human-oriented – abstractions
Programming progression...

• Programming has progressed through:
  – machine code;
  – assembly language;
  – machine-independent programming languages;
  – sub-routines;
  – procedures & functions;
  – abstract data types;
  – objects;

  to agents.
Global Computing

• What techniques might be needed to deal with systems composed of $10^{10}$ processors?
• Don’t be deterred by its seeming to be “science fiction”
• Hundreds of millions of people connected by email once seemed to be “science fiction”…
• Let’s assume that current software development models can’t handle this…
Where does it bring us?

• Delegation and Intelligence imply the need to build computer systems that can act effectively on our behalf

• This implies:
  – The ability of computer systems to act *independently*
  – The ability of computer systems to act in a way that *represents our best interests* while interacting with other humans or systems
Interconnection and Distribution

- Interconnection and Distribution have become core motifs in Computer Science.
- But Interconnection and Distribution, coupled with the need for systems to represent our best interests, implies systems that can cooperate and reach agreements (or even compete) with other systems that have different interests (much as we do with other people).
So Computer Science expands…

• These issues were not studied in Computer Science until recently
• All of these trends have led to the emergence of a new field in Computer Science: multiagent systems
Agents, a Definition

• An agent is a computer system that is capable of *independent* action on behalf of its user or owner (figuring out what needs to be done to satisfy design objectives, rather than constantly being told)
Multiagent Systems, a Definition

• A multiagent system is one that consists of a number of agents, which *interact* with one-another.

• In the most general case, agents will be acting on behalf of users with different goals and motivations.

• To successfully interact, they will require the ability to *cooperate*, *coordinate*, and *negotiate* with each other, much as people do.
Spacecraft Control

• When a space probe makes its long flight from Earth to the outer planets, a ground crew is usually required to continually track its progress, and decide how to deal with unexpected eventualities. This is costly and, if decisions are required *quickly*, it is simply not practicable. For these reasons, organizations like NASA are seriously investigating the possibility of making probes more autonomous — giving them richer decision making capabilities and responsibilities.

• *This is not fiction: NASA’s DS1 has done it!*
Deep Space 1

- “Deep Space 1 launched from Cape Canaveral on October 24, 1998. During a highly successful primary mission, it tested 12 advanced, high-risk technologies in space. In an extremely successful extended mission, it encountered comet Borrelly and returned the best images and other science data ever from a comet. During its fully successful hyperextended mission, it conducted further technology tests. The spacecraft was retired on December 18, 2001.” – NASA Web site
Internet Agents

• Searching the Internet for the answer to a specific query can be a long and tedious process. So, why not allow a computer program — an agent — do searches for us? The agent would typically be given a query that would require synthesizing pieces of information from various different Internet information sources. Failure would occur when a particular resource was unavailable, (perhaps due to network failure), or where results could not be obtained.
Agent Design, Society Design

• The course covers two key problems:
  – How do we build agents capable of independent, autonomous action, so that they can successfully carry out tasks we delegate to them?
  – How do we build agents that are capable of interacting (cooperating, coordinating, negotiating) with other agents in order to successfully carry out those delegated tasks, especially when the other agents cannot be assumed to share the same interests/goals?

• The first problem is agent design, the second is society design (micro/macro)
Multiagent Systems

In Multiagent Systems, we address questions such as:

- How can cooperation emerge in societies of self-interested agents?
- What kinds of languages can agents use to communicate?
- How can self-interested agents recognize conflict, and how can they (nevertheless) reach agreement?
- How can autonomous agents coordinate their activities so as to cooperatively achieve goals?
Multiagent Systems

• While these questions are all addressed in part by other disciplines (notably economics and social sciences), what makes the multiagent systems field unique is that it emphasizes that the agents in question are computational, information processing entities.
What is AI

Views of AI fall into four categories:

<table>
<thead>
<tr>
<th>Thinking humanly</th>
<th>Thinking rationally</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acting humanly</td>
<td>Acting rationally</td>
</tr>
</tbody>
</table>
Acting humanly: Turing Test

- Turing (1950) "Computing machinery and intelligence":
  - "Can machines think?" → "Can machines behave intelligently?"
  - Operational test for intelligent behavior: the Imitation Game

- Predicted that by 2000, a machine might have a 30% chance of fooling a lay person for 5 minutes
- Anticipated all major arguments against AI in following 50 years
- Suggested major components of AI: knowledge, reasoning, language understanding, learning
Thinking humanly: cognitive modeling

- 1960s "cognitive revolution": information-processing psychology
- Requires scientific theories of internal activities of the brain
- -- How to validate? Requires
  1) Predicting and testing behavior of human subjects (top-down)
     or 2) Direct identification from neurological data (bottom-up)
- Both approaches (roughly, Cognitive Science and Cognitive Neuroscience) are now distinct from AI
Thinking rationally: Laws of thought

- Aristotle: what are correct arguments/thought processes?
- Several Greek schools developed various forms of logic: notation and rules of derivation for thoughts; may or may not have proceeded to the idea of mechanization.
- Direct line through mathematics and philosophy to modern AI.
- Problems:
  1. Not all intelligent behavior is mediated by logical deliberation.
  2. What is the purpose of thinking? What thoughts should I have out of all the thoughts that I could have?
Acting rationally

- **Rational** behavior: doing the right thing
- The right thing: that which is expected to maximize goal achievement, given the available information
- Doesn't necessarily involve thinking – e.g., blinking reflex – but thinking should be in the service of rational action
What is an Agent?

- The main point/consensus about agents is they are autonomous (tự trị): capable of acting independently, exhibiting control over their internal state.
- One definition: *an agent is a computer system capable of autonomous action in some environment in order to meet its design objectives*.
What is an Agent?

- Trivial (non-interesting) agents:
  - thermostat
  - UNIX daemon (e.g., biff - email indicator)
Multi-agent System

• A multiagent system is one that consists of a number of agents, which *interact* with one-another.
• In the most general case, agents will be acting on behalf of users with different goals and motivations.
• To successfully interact, they will require the ability to *cooperate*, *coordinate*, and *negotiate* with each other, much as people do.
Environments – Accessible vs. inaccessible

• Tính truy cập được
• An accessible environment is one in which the agent can obtain complete, accurate, up-to-date information about the environment’s state
• Most moderately complex environments (including, for example, the everyday physical world and the Internet) are inaccessible
• The more accessible an environment is, the simpler it is to build agents to operate in it
Environments – Deterministic vs. non-deterministic

- Tính xác định
- A deterministic environment is one in which any action has a single guaranteed effect — there is no uncertainty about the state that will result from performing an action.
- The physical world can to all intents and purposes be regarded as non-deterministic.
- Non-deterministic environments present greater problems for the agent designer:
  - Agents have partial control over their environment.
  - Agents’ actions can fail.
Environments - *Episodic vs. non-episodic*

- Tính chia đoạn
- In an episodic environment, the performance of an agent is dependent on a number of discrete episodes, with no link between the performance of an agent in different scenarios
- Episodic environments are simpler from the agent developer’s perspective because the agent can decide what action to perform based only on the current episode — it need not reason about the interactions between this and future episodes
Environments - *Static vs. dynamic*

- Tính chất động và tĩnh
- A static environment is one that can be assumed to remain unchanged except by the performance of actions by the agent
- A dynamic environment is one that has other processes operating on it, and which hence changes in ways beyond the agent’s control
- Other processes can interfere with the agent’s actions (as in concurrent systems theory)
- The physical world is a highly dynamic environment
Environments – *Discrete vs. continuous*

- Tính rỗi rạc
- An environment is discrete if there are a fixed, finite number of actions and percepts in it
- Russell and Norvig give a chess game as an example of a discrete environment, and taxi driving as an example of a continuous one
- Continuous environments have a certain level of mismatch with computer systems
- Discrete environments could *in principle* be handled by a kind of “lookup table”
What is an Intelligent Agent?

• An intelligent agent is a computer system capable of flexible autonomous action in some environment

• By flexible, we mean:
  – Reactive (có phản ứng)
  – pro-active (chủ động có mục đích)
  – Social (có tính xã hội)
Reactivity

• If a program’s environment is guaranteed to be fixed, the program need never worry about its own success or failure – program just executes blindly
  – Example of fixed environment: compiler

• The real world is not like that: things change, information is incomplete. Many (most?) interesting environments are dynamic

• Software is hard to build for dynamic domains: program must take into account possibility of failure – ask itself whether it is worth executing!

• A reactive system is one that maintains an ongoing interaction with its environment, and responds to changes that occur in it (in time for the response to be useful)
Proactiveness

• Reacting to an environment is easy (e.g., stimulus → response rules)
• But we generally want agents to do things for us
• Hence goal directed behavior
• Pro-activeness = generating and attempting to achieve goals; not driven solely by events; taking the initiative
• Recognizing opportunities
Balancing Reactive and Goal-Oriented Behavior

• We want our agents to be reactive, responding to changing conditions in an appropriate (timely) fashion
• We want our agents to systematically work towards long-term goals
• These two considerations can be at odds with one another
• Designing an agent that can balance the two remains an open research problem
Social Ability

• The real world is a multi-agent environment: we cannot go around attempting to achieve goals without taking others into account!
• Some goals can only be achieved with the cooperation of others!
• Similarly for many computer environments: witness the Internet!
• Social ability in agents is the ability to interact with other agents (and possibly humans) via some kind of agent-communication language, and perhaps cooperate with others!
Other Properties

- Other properties, sometimes discussed in the context of agency:
  - **mobility**: the ability of an agent to move around an electronic network
  - **veracity**: an agent will not knowingly communicate false information
  - **benevolence**: agents do not have conflicting goals, and that every agent will therefore always try to do what is asked of it
  - **rationality**: agent will act in order to achieve its goals, and will not act in such a way as to prevent its goals being achieved — at least insofar as its beliefs permit
  - **learning/adaption**: agents improve performance over time