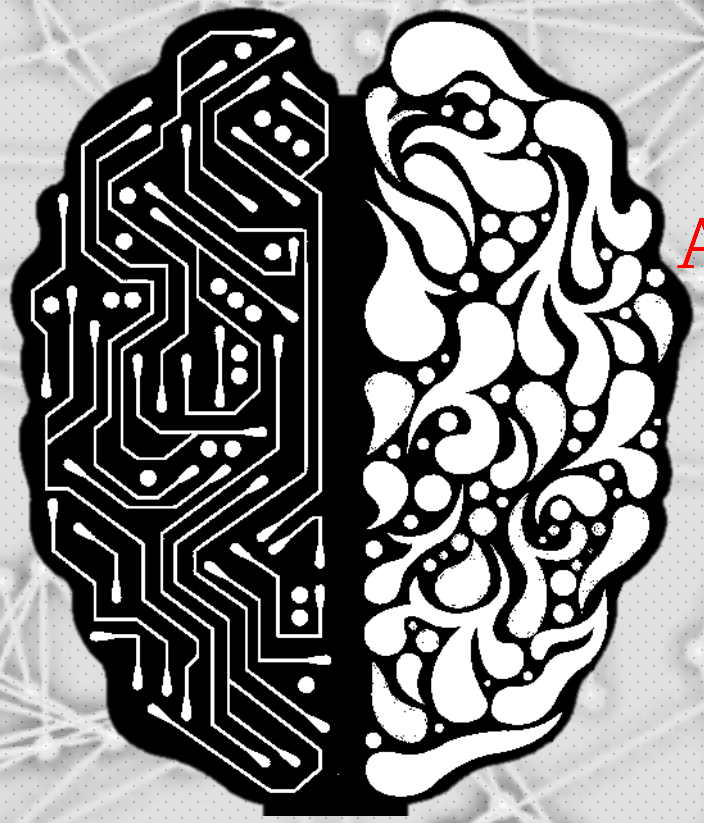


Multi-Agent Systems

Lecture II

- Dr. Nestor Velasco Bermeo,

- Researcher CONSUS (Crop OptimisatioN through Sensing, Understanding & viSualisation),
- School of Computer Science
- University College Dublin (UCD)

Multi-agent Systems Concepts

Distributed Artificial Intelligence - (DAI)

Traditionally, Artificial Intelligence has focused on how single human intelligence works.

- However, we do not act alone - a key feature of human society is our ability to communicate and cooperate.
- This led to the emergence, in the 1970s, of a subfield of AI research, known as Distributed AI.

- DAI is concerned with:
  - "*the development of distributed solutions for complex problems regarded as requiring intelligence*."

- Because of its aims and objectives, DAI research draws on a variety of fields:
  - Philosophy, Social Sciences, Economics / Game Theory, Linguistics, Computer Science/Engineering, …

Distributed Artificial Intelligence

- By the end of the 1980s, DAI research split:
  - **(Cooperative-) Distributed Problem Solving:**
    Designing networks of semi-autonomous processing nodes that work together to solve a given type of problem.
    - *Concerned with*: problem decomposition, task allocation, result synthesis, system optimisation
    - *Main technologies*: Distributed Constraint Programming / Optimisation.
  - **Multi-Agent Systems:**
    Understanding how groups of computational entities, known as agents, can collaborate and cooperate in order to solve problems that are beyond their individual capabilities.
    - *Concerned with*: intelligent decision-making, coordination, negotiation, organisation, distributed problem solving, software engineering.
    - *Main technologies*: anything goes!

Why Distributed Artificial Intelligence?

- Mirrors Human Cognition
- Potential Performance Enhancements
- Elegantly Reflects Society
- Incremental Development
- Increased Robustness
- Reflects Trends in Computer Science in General
- Strong Analogies to Decompositional Techniques employed in Software Engineering

Distributed Artificial Intelligence

- Endeavours to achieve Intelligent Systems not by constructing a large Knowledge-Based System, but rather by partitioning the knowledge domain and developing *'Intelligent Agents'*, each exhibiting expertise in a particular domain fragment.

- This group of agents will thereafter collectively work towards the solution of global problems.

Problems with DAI

- Identification of appropriate **task decomposition** and **task distribution strategies**.
- Optimisation of problem solution (Cammarata et al 1982,1983)
- Difference of opinion between experts where the mapping between expertise and experts is not 1::1 but 1::n - need conflict resolution strategies
- Problems with understanding
- Handling uncertainty
- Deadlock avoidance strategies
- Heterogenous nodes
- Interoperability

Multi-Agent Research Topics

Theories of Agency

- Logical Models of Rational Action
- Game Theoretical Approaches
- Planning

Agent-Oriented Software Engineering

- Tools, Languages and Methodologies
- Environments
- Standards

Multi-Agent Research Topics

Multi-Agent Interaction

- Cooperation and Coordination
- Organisations & Institutions
- Negotiation
- Distributed Planning

Multi-Agent Learning & Problem Solving

The Co-operating Experts Metaphor

This solution of problems by a group of agents, providing mutual assistance and when necessary is often referred to as the...

"*Community of Co-operating Experts Metaphor*"

Smith and Davis, Lenat, Hewitt

Proponents of this philosophy believe that reciprocal co-operation is the cornerstone of society.

Agents are Embodied AI

- (Russell and Norvig, 1995) state that an **agent** is:
  - *"anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators"*

- Thus, they view an agent as:
  - any entity that is *located* in some environment, and which
  - interacts with that environment through a set of sensors and actuators.

- They then extend this definition to identify an **intelligent agent** as any agent that **embodies some AI technique**.
  - This does not just apply to Expert Systems, but also to machine learning algorithms, planning algorithms, …

The Great Agent Debate

- The term "**agent**" means different things to different people.

"An agent is a computer system that is situated in some environment, and that is capable of flexible, autonomous action in this environment in order to meet its design objectives"

(Wooldridge and Jennings, 1995)

How others define Agents…

- MuBot Agent
- AIMA Agent
- Maes Agent
- KidSim Agent
- Hayes-Roth Agent
- IBM Agent
- Wooldridge & Jennings Agent
- SodaBot Agent
- Foner Agent
- Brustoloni Agent

How others define Agents…

- **MuBot Agent;** autonomous execution & ability to perform domain oriented reasoning.
- **AIMA Agent;** anything that can perceive its environment through sensors and acting through effectors.
- **Maes Agent;** inhabit complex dynamic environment and act autonomously realizing a set of goals/tasks for which they were designed.
- **KidSim Agent;** persistent software entity dedicated to a specific purpose.
- **Hayes-Roth Agent;** perceive dynamic environment, act to affect it and reason to interpret perception, solve problems, draw inferences and determine actions.

How others define Agents…

- **IBM Agent;** software entities that carry out a set of operations with a degree of autonomy/independence employing knowledge or the user's goals/desires.
- **Wooldridge & Jennings Agent;** agent expose 4 key properties: autonomy, social ability, reactivity and pro-activeness.
- **SodaBot Agent;** programs that engage in dialogs, negotiate and coordinate information transfer.
- **Foner Agent;** collaborate to accomplish user's task while being autonomous , trustworthy and degrade gracefully to a communication mismatch.
- **Brustoloni Agent;** capable of autonomous, purposeful actions in the real world.

The Great Agent Debate…

- In contrast, (Maes, 1995) views agents to be:

  "computational systems that inhabit some complex dynamic environment, sense and act autonomously in this environment, and by doing so realise a set of goals or tasks for which they are designed."

- This posits a view of an agent as:
  - any autonomous software entity that is located in a **complex dynamic environment**, and which
  - exhibits **goal-oriented behaviour**, requiring that it act in pursuit of its own goals.

The Great Agent Debate…

- Alternatively, (Shoham, 1993) adopts the perspective that:

  "An agent is an entity whose state is viewed as consisting of mental components such as beliefs, capabilities, choices, and commitments. These components are defined in a precise fashion, and stand in rough correspondence to their common sense counterparts"

- This third definition adopts the view of agents as **mental entities**:
  - That is, entities that employ mental concepts such as beliefs, commitments, and goals in order to reason about both the environment and their activities…

One Definition to Rule them All…

- In 1995, Michael Wooldridge and Nick Jennings proposed a **two-tier definition** of agency that has become a de facto standard for agent research.
- The lower tier, or **weak notion of agency**, was intended to be sufficiently general to meet the needs of most agent researchers, and specified the following agent attributes:
  - **Autonomy, *social ability*, reactivity, and pro-activity**.
- The upper tier, or **stronger notions of agency**, were intended to build on this weak core to provide more specific definitions, and specified attributes such as:
  - **Benevolence, rationality, mobility, learning, intentionality**, …

Weak Agency

- **Autonomy:** Agents operate without the direct intervention of humans or others, and have some kind of control over their actions and internal state.

- **Social Ability:** Agents interact with other agents and (possibly) humans via some kind of agent communication language.

- **Reactivity:** Agents perceive their environment (which may be the physical world, a user via a graphical user interface, a collection of other agents, the Internet, or perhaps all of these combined), and respond in a timely fashion to changes that occur in it.

- **Pro-activity:** Agents do not simply act in response to their environment, they are able to exhibit goal-directed behaviour by taking the initiative

Strong Agency

- **Mobility:** the ability of an agent to move around an electronic network.

- **Benevolence:** Is the assumption that agents do not have conflicting goals, and that every agent will therefore always try to do what is asked of it.

- **Rationality:** is (crudely) the assumption that an 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 in so far as its beliefs permit.

- **Intentionality:** an agent reasons about its activities through the application of mental notions such as beliefs, goals, obligations, commitments, intentions…

Reactive VS Intentional Systems

- Essentially Multi-Agent systems occupy a point on a continuum between two extreme classes of system. These two extremes are...

The classical system

The reactive or situated action system

- We propose a compromise that of the

'Deliberate Social Agent'

Reactive or Situated Systems

- Agents **react** to varying situations and consequently **do not** have an **explicit** representation of the world within which they exist.

- **Reasoning** takes place within each agent at a very **low level**, essentially each agent has little more than an ability to perform **pattern matching**.

- A given **situation** is characterised and matched against a **collection of rules** specifying appropriate **behaviour** associated with each of these situations

i.e. situation → action or situated action.

Reactive or Situated Systems

- Typically the actions associated with a given situation are often very simple and consequently the agents themselves are very simple computational entities.
- Even though each of the individual agents are very simple the global complexity and global structures can be achieved as a result of the emergent property of the interacting behaviours of the community of agents.

Reactive Systems Assessment

Advantages

- simplicity.
- avoidance of necessity for a sophisticated representation of the world and more significantly the problems of maintaining this model.
- generally the structure of agent interaction is well defined and domain independent.

Disadvantages

- New sets of rules need to be designed for each application.
- Each situation needs to be specified and identified so as to have an associated rule.
- Difficulty in solving inherently recursive problems.
- Lack of a precise theory upon which the combining behaviours of agents can be based and explained.

Intentional Systems

- Generally the agents within a reflective system are more complex computational entities.

- They do not merely react to a given situation in a specific way. In fact they may often react in different ways dependent on their own ‘beliefs’ or ‘intentions’.

- Such systems necessitate an internal representation of the world. They often base their reasoning on the actions of the other agents within the community.

Intentional Systems

- They normally possess some model of intentionality which represents their goals, desires, prejudices, beliefs etc. about themselves and the remainder of the community.

- Certain classes of problem seem to necessitate this ability to reason using intentionality. The ‘wisest man’ puzzle seems to typify these.

Intentional Systems

- Reasoning intentionally normally demands use of **higher order logics**.

- Modal logics.

  - Epistemic logics
  - Doxastic logics

- There are two general approaches

  - Sentential logics (Konolidge)
  - Possible World Logics (Kripke)

The Intentional Stance

- In arriving at the philosophy of intentional systems (Dennett, 1989) draws heavily on what he calls folk psychology which he defines as:
  - *a perspective which invokes the family of "mentalistic" concepts, such as belief, desire, knowledge, fear, pain, expectation, intention...*

- Based on the view that human behaviour is often explained using these mentalistic concepts:
  - e.g. "Joe hit Bill because he wanted his bike".

The Intentional Stance

- This view of decision-making is inspired by the work of the philosopher Daniel Dennett (1989) who identifies 3 levels at which behaviour can be modelled:
  - **Physical Stance**: the domain of physics and chemistry; concerned with mass, energy, velocity, chemical composition, …
    - *Predicting where a ball will land based on trajectory.*
  - **Design Stance**: the domain of biology and engineering; concerned with purpose, function and design.
    - *Predicting that a bird will fly when flapping its wings because this is what wings are for.*
  - **Intentional Stance**: the domain of software and minds; concerned with belief, thinking, and intent.
    - *Predicting that the bird will fly away because it knows the cat is coming and it is afraid of being eaten.*

Intentional Stance and Agents

- Using the Intentional Stance allows:
  - Abstraction from the underlying system complexity
    - Beliefs and knowledge, wants and desires, fears and joys, …
  - Simple to model rational decision-making processes:
    - X intends to move away from Y because X believes Y is too close and is afraid of Y.
    - The robot goes to the fridge because it believes that its master wants a beer.
  - Sits well with logic:


```
Believes(X, close(Y)) & Afraid(X, Y) =>
     Intends(X, moveFrom(Y))
Believes(robot, wants(master, beer)) =>
     Intends(robot,
          goto(fridge);get(beer);goto(master);give(beer))
```

Intentional Stance and Agents

- The Argument:

  Viewing the behaviour of software systems from an intentional stance allows us to provide a more abstract (simpler) definition of that behaviour. This, in turn, allows us to build more complex software...

- Some argue that the use of the Intentional Stance is a pointless attempt to anthropomorphise programming.
  - "A fancy lookup table"
  - "An unnecessarily overcomplicated programming paradigm"
  - "What is the benefit of mental state programming?"

Intentional Stance and Agents

- While several “mental models” have been proposed, a *de facto* standard, known as the **Belief-Desire-Intention (BDI)** architecture, has emerged:
  - **Beliefs**: the current state of the environment
  - **Desires**: the agent ideal future state of the environment
  - **Intentions**: subset of the desires that the agent commits to

Intentional Stance and Agents

- Informally, **BDI** theories attempt to capture the transition between states.
  - Desires drive the agent's activities and are satisfied when the agent believes that it has achieved them.
  - Agents are resource bounded: desires may be incompatible.
  - Intentions represent the trade off that the agent makes in terms of the subset of its desires that it commits to achieving.

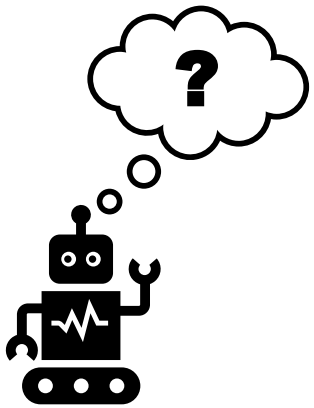
Is the BDI model now becoming somewhat dated? Why?

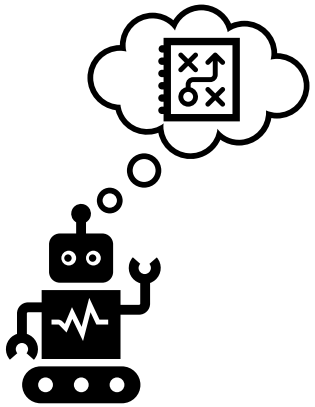
Example

While true

1. Observe the world
2. Update internal world model
3. *Deliberate about what intention to achieve next*
4. *Use means end reasoning to get a plan for the intention*
5. Execute plan

End while





Lecture II Learning Objectives (recap)

- An Expert System (ES) consists of a Database, Inference Engine and a Rule Base.
- An ES focuses on a problem domain and a formal representation of expertise knowledge.
- Two inference engines for ES: Forwards Chaining and Backwards Chaining.
- DAI focuses on distributed solving complex problems that require intelligence.
- MAS occupy a point on a continuum between two extreme classes of systems.
- There are several definitions for what an Agent is.
- There are two types of agency (Strong and Weak).

Things to Do!

- Look at relevant chapters from:

Wooldridge, M. (2009). An introduction to multiagent systems. John Wiley & Sons.

- Augment Notes from the following paper:

Wooldridge, M., & Jennings, N. R. (1995). Intelligent agents: Theory and practice. The knowledge engineering review, 10(2), 115-152.

- Supplement notes from Chapter 1 of:

O'Hare, G. M., Jennings, N. R., & Jennings, N. (Eds.). (1996). *Foundations of distributed artificial intelligence* (Vol. 9). John Wiley & Sons.