EECS 3401 — AI and Logic Prog. — Lecture 1

Adapted from slides of Prof. Yves Lesperance

York University

September 14, 2020

- EECS 3401: "Introduction to Artificial Intelligence and Logic Programming"
- Instructor: Vitaliy Batusov (contact: `vbatusov@cse.yorku.ca`)
- Course textbook: Russell & Norvig, *Artificial Intelligence: A Modern Approach*, 4th edition (2020).
- Lecture schedule: Monday & Wednesday, 14:30–16:00 on Zoom
- Office Hours: TBA soon, check eClass

Syllabus

Will cover fundamental concepts of AI:

- intelligent agents
- knowledge representation and reasoning — FOL
- search (uninformed, informed)
- constraint satisfaction, backtracking
- reasoning about action; planning
- reasoning under uncertainty — Bayesian Networks
- logic programming — Prolog

Evaluation

- 3 assignments ($8\% \times 3 = 24\%$)
- Midterm (26%)
- Exam (50%)

- AI = Artificial Intelligence
- What is intelligence?

  Something along the lines of *the capacity to acquire and apply knowledge*, *the faculty of thought and reason*
- What features/abilities/behaviours are indicative of intelligence?
- Has to do with deliberate action in a wide variety of circumstances

Variety among Definitions

As per Russell & Norvig, book definitions of intelligent systems broadly fall into one of the categories:

Think like humans	Think rationally
Act like humans	Act rationally

Turing test

Human interrogator communicates with hidden subject; must decide whether subject is a **human** or a **machine**. If human can't reliably identify the machine, the machine passes the test.

- Highly influential definition
- Good reasons to consider a system that passes the test intelligent
- No insight on how to build such a machine

Human Intelligence

So how do we build AI?

- Let's imitate natural (human) intelligence
  - It exists
  - It works
  - It can be observed and studied

Human Intelligence

Human intelligence is built on fundamentally different hardware:

- Biological vs. electronic
- Vast disparity re: numerical computations
- Visual and sensory processing
- Massive-yet-slow parallel vs. lightning-fast serial processing

Also, built by a fundamentally different process.

Human Intelligence

- Very hard to look under the hood of human intelligence.
- Little is known about the high-level processing in the brain; hard to replicate something you have no scientific understanding of.
- Nevertheless, neuroscience has been influential in some areas (robotic sensing, computer vision, etc.)

Rationality

- Human intelligence can't be said to be perfectly rational
- **Rationality**: a precise mathematical notion of what it means to do the right thing in any particular circumstance
- A precise mechanism for analyzing and understanding properties of the ideal behaviour we are trying to achieve
- A precise benchmark against which to measure the performance of systems we build

Rationality

- Mathematical characterizations of rationality have come from diverse areas
- Logic — laws of reasoning
- Economics — utility theory, acting under uncertainty, game theory
- No agreement about which notion of rationality is best
- Not that important as long as they are precise
- This course: **acting rationally**

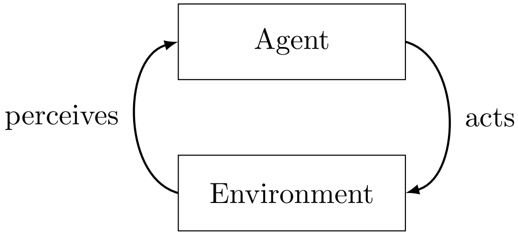
Computational Intelligence

- *AI tries to understand and model intelligence as a computational process*
- Try to construct systems whose **computation** achieves or approximates the desired notion of rationality
- Hence, AI is part of Computer Science

Agency

- It is useful to think of intelligent systems as being *agents* with own goals or acting on behalf of someone else
- An **agent** is an entity that exists in an **environment** and that **acts** on said environment based on its **perceptions** of the environment.
- An **intelligent agent** acts to further its own interests (or those of a user)
- An **autonomous agent** can make decisions without user's intervention, possibly based on its own learning

Agent and Environment

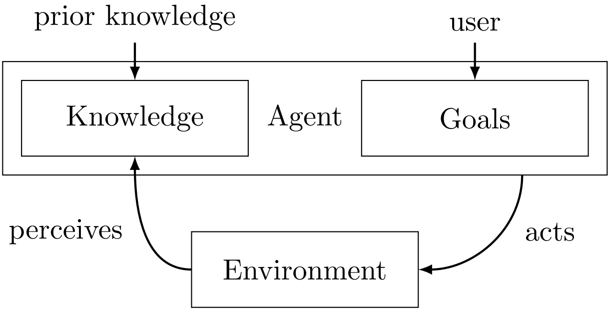


Note: this diagram ignores the internal structure of the agent

Types of agents

- **Simple reflex agents**: apply simple condition-action rules to decide next action based on current percepts
- **Model-based reflex agents**: maintain a *model* of the world, apply rules to decide next action based on current world model
- **Goal-based agents**: decide next action based on current model of the world state and current *goal(s)*; may do planning, more flexible

A better agent



This agent supports more flexible interaction with the environment, can modify its goals, and can flexibly apply its knowledge to different situations

Types of agents (cont.)

- **Utility-based agents**: choose actions to maximize their *expected utility* in *uncertain* worlds

All types of agents can benefit from a *learning mechanism*: explore space of possible rules/actions/models, evaluate performance, and modify agent to improve and adapt

Environments

- Fully observable vs. Partially observable
- Deterministic vs. Stochastic
- Episodic vs. Sequential
- Static vs. Dynamic
- Discrete vs. Continuous
- Single-agent vs. Multi-agent
- Known dynamics vs. Unknown dynamics

Agent Architectures

- Agents may have more complex architecture than we've seen so far
- Embodied agents (e.g., robots) tend to have complex hierarchical control architectures with multiple layers
  - Low-level: local motion and collision avoidance
  - Mid-level: path planning and following
  - High-level: task planning

Degrees of Intelligence

- Human-level AI remains an elusive goal
- Local successes in specialized forms of intelligence
- Useful formalisms and algorithms for “intelligent systems” have been developed
- These form the foundation for our attempt to understand intelligence as a *computational* process

- In this course, we will study some of these formalisms and see how they can be used to achieve various degrees of intelligence

Hall of Fame

- 1997 IBM Deep Blue beats world chess champion
- 1999 NASA Remote Agent uses AI planning to control spacecraft
  Autonomy becomes routine in robotic missions to planets
- 2005 5 robot cars complete 212-km course through Mojave desert
  DARPA Grand Challenge
- 2011 IBM Watson beats best humans in Jeopardy
  When asked a tricky question about US cities, Watson answered "Toronto"[1]
- 2016 DeepMind AlphaGo beats best human in Go
- 2019 Tesla cars autonomously navigate parking lots — an extremely open and challenging environment[2]
- "soon" A feature-complete self-driving Tesla

[1] https://www.youtube.com/watch?v=7h4baBEi0iA

[2] Like Watson, it's not without issue
https://twitter.com/eiddor/status/1177749574976462848

What's behind recent progress

- Overall better hardware
- In ML, dedicated highly-parallelized computing
- Improving techniques
  - Better search methods and heuristics
  - Better representations
  - Availability of large datasets

Sub-areas of AI

- Perception: computer vision, speech understanding
- Robotics
- Natural language understanding
- Machine learning
- Reasoning and decision making *(you are here)*
  - **Knowledge representation**
  - **Reasoning** (logical, probabilistic)
  - **Decision making** (search, planning, decision theory)

Prospects

- Will rapid progress continue?
- Concerns about risks of developing AI
  - Robots enslaving humans — probably not
  - Humans using AI as a weapon — you bet
- Are current learning-based AI systems really intelligent?

Winograd Schema Challenge: resolving the ambiguity using *common sense*

The city councilmen refused the demonstrators a permit because they **feared** *violence* — who feared violence?

The city councilmen refused the demonstrators a permit because they **advocated** *violence* — who advocated violence?

End of lecture

Next time: Knowledge Representation & First-Order Logic