Artificial Intelligence Fundamentals and Intelligent Agents

Danilas Miscenko

September 1, 2020

Acknowledgements

While working on this assignment, some topics were discussed in a group. The group members are Piri Babayev, Danilas Miscenko (me) and Aleksander Simmersholm. Because of the discussion that has taken place, some similarities between answers may arrise, but all of the assignment tasks were completed individually, as the point of the discussions is to get a better understanding of the subject and not copy each others work.

1 Define artificial intelligence (AI). Find at least 3 definitions of AI that are not covered in the lecture.

Three definitions found online (Marsden, 2017):

Artifitial Intelligence is...

- "... is the next, logical step in computing: a program that can figure out things for itself. It's a program that can reprogram itself" (Jim Sterne, author of Artificial Intelligence for Marketing)
- "... the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings" (definition in Encyclopedia Britannica by Prof. B.J. Copeland)
- "... a computerized system that exhibits behavior that is commonly thought of as requiring intelligence" US Government definition (NSTC)

2 What is the Turing test, and how it is conducted?

The curriculum book states that "the Turing test was designed to provide a satisfactory operational definition of intelligence" (Russel et al., 2010). The test is conducted by having a human interrogator pose some questions through some sort of chat function. The interrogator has no knowledge on whether the one answering the questions is also human or an AI. The test is passed if the interrogator cannot tell whether the posed questions were answered by a human or an AI.

3 What is the relationship between thinking rationally and acting rationally? Is rational thinking an absolute condition for acting rationally?

To think rationally is to infer the best action to achieve some sort of goal, while to act rationally is to preform said action.

Rational thought can sometimes take too long to compute however. An example of this can be seen in humans when putting your hand on a hot stove. The rational action is to remove the hand from the stove, but it takes too long for the brain to deliberate on removing the hand before sending the signals to the muscles. The human body has therefore evolved reflexes, which bypass the brain entirely by sending the signals to the muscles as soon as pain is registered. This example shows that thinking rationally is not an absolute condition for acting rationally.

4 Describe rationality. How is it defined?

The curriculum book says, "a system is rational if it does the "right thing", given what it knows" (Russel et al., 2010). The "right thing" mentioned in the quote is referring to some sort of action meant to achieve a specified goal. Given this definition, rationality can be further described as a performance measure of how effective any given action is to achieve a set goal, or in other words, rationality maximises expected performance.

5 Some AI history

a) What is Aristotle's argument about the connection between knowledge and action? Does he make any further suggestion that could be used to implement his idea in AI?

The following is an excerpt from the curriculum book used in the course (Russel et al., 2010).

«Aristotle argued that actions are justified by a logical connection between goals and knowledge of the action's outcome:

```
But how does it happen that thinking is sometimes
accompanied by action and sometimes not, sometimes by
motion, and sometimes not? It looks as if almost the
same thing happens as in the case of reasoning and
making inferences about unchanging objects. But in
that case the end is a speculative proposition ...
whereas here the conclusion which results from the two
premises is an action. ... I need covering; a cloak is
a covering. I need a cloak. What i need, I have to make;
I need a cloak. I have to make a cloak. And the conclusion,
the "I have to make a cloak," is an action.
```

In the *Nicomachean Ethics*, Aristotle further elaborates on this topic suggesting an algorithm:

```
We deliberate not about ends, but about means. For a
doctor does not deliberate whether he shall heal, nor
an orator whether he shall persuade, ... They assume
the end and consider how and by what means it is
attained, and if it seems easily and best produced
thereby; while if it is achieved by one means only
they consider how it will be achieved by
```

```
this and what means *this* will be achieved,
till they come to the first cause, ... and what is
last in the order of analysis seems to be first in the
order of becoming. And if we come on an impossibility,
we give up on the search, e.g., if we need money and
this cannot be got; but if a thing appears possible we
try to do it. »
```

b) Who was/were the first AI researcher(s) to implement these ideas?

The first researchers to implement this algorithm were Herbert A. Simon, J. C. Shaw, and Allen Newell.

c) What is the name of the program/system they developed?

The system was called the General Problem Solver, or GPS for short.

d) Google about this system and write a short description about it

Excerpt from the Wikipedia article about GPS (Wikipedia, 2020):

"It was based on Simon and Newell's theoretical work on logic machines. GPS was the first computer program which separated its knowledge of problems (rules represented as input data) from its strategy of how to solve problems (a generic solver engine). GPS was implemented in the third-order programming language, IPL.

While GPS solved simple problems such as the Towers of Hanoi that could be sufficiently formalized, it could not solve any real-world problems because search was easily lost in the combinatorial explosion. Put another way, the number of "walks" through the inferential digraph became computationally untenable. ...

The user defined objects and operations that could be done on the objects, and GPS generated heuristics by means-ends analysis in order to solve problems. It focused on the available operations, finding what inputs were acceptable and what outputs were generated. It then created subgoals to get closer and closer to the goal."

6 Consider a robot whose task it is to cross the road. Its action portfolio looks like this: look-back, look-forward, look-left-look-right, go-forward, go-back, go-left and go-right.

a) While crossing the road, a helicopter falls down on the robot and smashes it. Is the robot rational?

Since the robot is not able to look upwards, it couldn't have perceived the helicopter crashing onto it. This means that the robot was acting rationally, as it couldn't predict being crashed upon.

b) While crossing the road on a green light, a passing car crashes into the robot, preventing it from crossing. Is the robot rational?

Here, the robot is not rational. Because it is able to perceive cars, it should have seen the car coming and got out of harms way.

7 Consider the vacuum cleaner world described in Chapter 2.2.1 of the textbook. Let us modify this vacuum environment so that the agent is penalized 1 point for each movement.

a) Can a simple reflex agent be rational for this environment? Explain your answer.

A simple reflex agent would not be rational for such an environment because once the agent is done cleaning, it would oscillate back and fourth between the rooms, moving needlessly and accumulating penalty points for each movement.

b) Can a reflex agent with state be rational in this environment? Explain your answer.

By having state, the agent effectively gains memory. It can then remember which tiles it has cleaned, and thus does not need to check the tiles as often. It therefore would be rational, as it could just stop and not oscillate between the tiles, thereby minimizing the penalty points for movement.

c) Assume now that the simple reflex agent (i.e., no internal state) can perceive the clean/dirty status of both locations at the same time. Can this agent be rational? Explain your answer. In case it can be rational, design the agent function.

As the agent in this context is able to observe the environment fully, it doesn't need state to be rational. Since it knows that all tiles are clean, it doesn't need to move to check. Therefore it would just wait until one or both of the tiles get dirty before going in to clean them. This agent is therefore rational. The agent function would look something like this:

```
function Sucky-Vacuum-Cleaner-Agent(percept) returns an action
  persistent: percepts, information about the state of all tiles

  action <- LookUp(percepts)
  return action
```

8 Consider the vacuum cleaner environment shown in Figure 2.3 in the textbook. Describe the environment using properties from Chapter 2.3.2, e.g. episodic/sequential, deterministic/stochastic etc. Explain selected values for properties in regards to the vacuum cleaner environment.

Since the agent is not able to observe the whole environment, as it only has sensors allowing it to observe the tile it is in now, the environment is only **partially observable**.

Since the environment is meant for a **single agent**, it is neither **competitive** nor **cooperative**.

The state of one tile does not depend on whether another tile is dirty or not, which means that the environment is **stochastic**.

Due to the sensors the agent has access to, each tile can be seen as atomic. The act of perceiving whether the tile is clean or not, cleaning it and moving to the next tile is therefore its own separate episode, making the environment **episodic**.

The environment is **dynamic** because dirt on the tiles accumulates over time.

Even though dirt is continuously accumulating on the tiles, making cleaning the tiles an infinitely long task, the boolean states of the tiles, being either *dirty* or *clean*, and the small number of actions the agent has at its disposal makes it a **discrete** environment.

As the agents rules are very simple, as in **if** dirty **then** suck, **if** clean **then** move on, the rules of the environment also become very simple, making the environment **known**.

9 Discuss the advantages and limitations of these four basic kinds of agents:

a) Simple reflex agents

The simple reflex agents are, as the name suggests, simple. They react to their percepts and choose actions only based on what the world is like at the moment of choice. These kinds of agents are very good for actions that do not require much deliberation, but fair pretty badly compared to other kinds of agents when presented with more complex tasks.

b) Model-based reflex agents

A model-based reflex agent is an improved version of the simple reflex agent, and has state. This effectively gives it memory, which in turn allows it to choose an action based on previous choices. While this newfound reasoning is a big advantage, the agent does not deliberate on the outcome of its actions, meaning that it can become predictable when acting rationally.

c) Goal-based agents

The goal-based agent, as the name suggests, is given a goal to strive to, as well as a notion that achieving said goal will make it "happy". This in practise means that the agent deliberates on the choices it can take and how those choices impact the environment, before choosing the choice that brings it the closest to completing the goal. A drawback of this type of agent is that there is no gradient measurement of how well the goal is achieved, it's either achieved or not. This results in the agent being prone to take the actions that bring it closest to goal completion in the moment, but not consider the actions that might be less effective in the moment, but lead to the goal being achieved at a more satisfactory level.

d) Utility-based agents

The utility-based agent, just like the goal-based agent, is given a goal to complete, but differing from the goal-based agent, it quantifies it's performance.

What this means is that the agent has different parameters that measure how well the goal is being completed. This means that the agent is able to deliberate on how to complete its goals most efficiently, while also being able to learn from previous mistakes, and try to better its ability to complete said goals. A drawback of this type of agent is that it is computationally difficult. The time it takes for the agent to choose an action might be too great to be able to achieve rationality.

Bibliography

Marsden, P. (2017). Artificial intelligence defined: Useful list of popular definitions from business and science. *digitalwellbeing.org*. https://digitalwellbeing.org/artificial-intelligence-defined-useful-list-of-popular-definitions-from-business-and-science/.

Russel, S. J., Norvig, P., & Davis, E. (2010). *Artificial intelligence a modern approach* (3rd ed.). Upper Saddle River, NJ : Prentice Hall, Pearson Education, Inc.

Wikipedia. (2020). General Problem Solver — Wikipedia, the free encyclopedia [[Online; accessed 01-September-2020]].