

Assignment 1

Benjamin Sorenson

September 23, 2015

1 Written Questions

1. [20 points] Performance measures for an agent could be designed based on the effects the agent has on the environment or according to the behaviors of the agent. Explain briefly what is the difference, provide an example for each, and explain if one of the two choices is best and why.

Answer: Taking two examples from the book, in the case of the vacuum cleaner agent, one could define a performance measure by the amount of dirt cleaned up in a single 8 hour shift, awarding more points for more dirt cleaned—a performance measure according to the behaviours of the agent. Alternatively, one could define a performance measure by awarding one point for each clean square at each time step—a performance measure according to the effects the agent has on the environment.

In the case of a performance measure designed according to the behaviours of the agent, we reward behavior that we think will lead to the desired effects on the environment rather than reward for the desired effects directly. This can allow a rational agent to maximize its performance measure without the desired effects to the environment. For example, as the book points out, a rational vacuum cleaner agent that is maximizing the amount of dirt cleaned in an 8 hour period could maximize its performance measure by cleaning up the dirt, dumping it back on the floor and cleaning it again. For this reason, it is better to define a performance measure based on the desired effects to the environment rather than the behaviors of the agent.

2. [30 points] You are given the following problem: Given a 5-gallon jug filled with water and an empty 2-gallon jug how can you have precisely 1 gallon of water in the 2-gallon jug? Assume you can fill the jugs with water as many times as desired, but you cannot measure how much water is in each jug. When you move water out of a jug you can either fill up the other jug or dump the water.

You are to formulate the problem using a state-space search representation.

Describe (precisely):

- (a) what is the initial state

The initial state is a 5-gallon jug (referred to as “5G” for the remained or the description) filled with water and an empty 2-gallon jug (referred to as “2G” for the remainder of

the description). Note: for the remainder of this description a state will be denoted as $\{Jug1 : NumberofGallons, Jug2 : NumberofGallons\}$. For example, the initial state would be $\{5G : 5, 2G : 0\}$.

- (b) the goal test

Precisely 1 gallon of water in the 2G.

- (c) the actions (called in the textbook successor function)

- i. `pour_into(PouringJug, ReceivingJug)`

Definition: if the amount of water in `PouringJug` ($g1$) plus the amount of water in `ReceivingJug` ($g2$) is \leq to the capacity of `ReceivingJug` ($G2$), the amount of water in `ReceivingJug` is set to $g1 + g2$, otherwise, fill `ReceivingJug` to $G2$ and set the level of `PouringJug` to $g1 + g2 - G2$

Example:

$$RESULT(\{5G : 5, 2G : 0\}, \text{pour_into}(5G, 2G)) = \{5G : 3, 2G : 2\}$$

- ii. `empty(Jug)`

Definition: set the level of `Jug` to 0

Example:

$$RESULT(\{5G : 5, 2G : 0\}, \text{empty}(5G)) = \{5G : 0, 2G : 0\}$$

- iii. `fill(Jug)`

Definition: set the level of `Jug` to its maximum capacity

Example:

$$RESULT(\{5G : 5, 2G : 0\}, \text{fill}(2G)) = \{5G : 5, 2G : 2\}$$

- (d) the path cost

Each step costs 1 so the path cost is the number of steps in the path

- (e) the state-space for the problem

The state-space is any state of fullness of the two-gallon and five-gallon jugs reachable by the actions `pour_into(Jug1, Jug2)`, `fill(Jug)`, and `empty(Jug)` performed on the initial state and any resultant states. Though other states are possible, any state can be trivially converted to one of the states reachable by the initial state in this problem definition in at most two actions.

- (f) is the state-space a tree or a graph?

The state space is a graph since there are repeated states.

- (g) what search algorithm would you use and why?

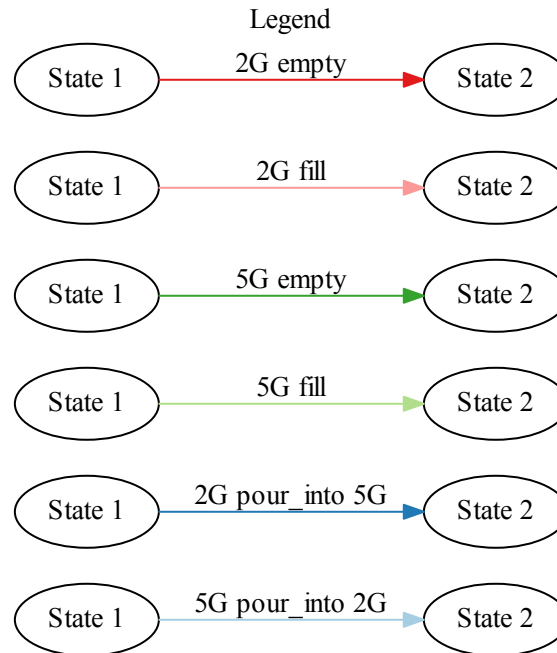
I would use a uniform-cost search because the search space is small, and it will always find an optimal solution.

- (h) show graphically the search space explored and the solution (there might be more than one solution)

Next last two pages of this document show

- i. The search space explored
- ii. The optimal solution

The color of the arrows indicate the action according to the following legend



3. [20 points] There has been a lot of discussion recently in the news and on the web on the dangers of AI, started by Stephen Hawking and Elon Musk

- (a) Search for a few on line writings on the controversy and summarize (in 1/2 to 1 page) the main arguments made.
- (b) Read the white paper by Eric Horvitz outlining the project One Hundred Year Study on Artificial Intelligence. Can you think of some additional topics not listed there? or could you add more questions to one of the topics listed? Summarize your thoughts (in a few paragraphs).

1.1 One Hundred Year Study on Artificial Intelligence

I wish I could spend more time thinking about the topics covered. In each topic, the central concern seems to be to one degree or another “What are the consequences of turning over traditionally human decisions over to machines?” and, in the other direction “What can we accomplish (for good for bad) with the aid intelligent machines?” I will focus on the areas of ethics and education—ethics because ethical issues seem to arise in even the most benign-sound goals (e.g. auto-driving vehicles), and education because I work in the education industry.

1.1.1 Ethics

Personally, I'm most looking forward to autonomous assistants that don't take any direction—they observe your behaviour and try to help out. For the first few weeks or months, they may do almost nothing, and suddenly rudimentary tasks begin to get completed on their own. Of course, this comes with its own set of ethical concerns. What if you're a serial killer, and you now have a serial killer assistant—even if checks are put in place so that the machine can't assist killing or some other hard safe-guard, real life is fuzzy. Where does the threshold lie between contributing to nefarious activity and not? Who decides?

1.1.2 Educations

I saw an info-graphic (I can't remember where, now), that put education near the bottom of fields most likely impacted by "Data Science." I tend to instead agree with the idea that education is one of the key areas of opportunity for AI. I've heard (source?) that the most difficult part of teaching is figuring out what when wrong when a student comes up with an incorrect solution—this difficulty is then multiplied by the number of students that a teacher has to observe. A machine that could perfectly and objectively observed, memorize, and perform rigorous analyses to assist a teacher in discovering these paths leading to incorrect understanding of the material, I would have to believe, would be invaluable to both teachers and students. Maybe this would take the form of a study tool where not only is the student shown the correct answer when, but also given instruction specific to the deficiency in their understanding of the material.

1.2 Why We Really Should Ban Autonomous Weapons: A Response

Why We Really Should Ban Autonomous Weapons: A Response Addresses four argument the authors see as being put forth by the AI community against a ban on autonomous weapons. The first addressed is that, and I'm paraphrasing, autonomous weapons are an inevitability so it makes no sense to try and ban them. The second is addressed along with the first, and it is that (again, paraphrasing) that the real problem is the willingness of humans to use technology for evil. These arguments, the authors say, apply to *any* weapons system, but we still ban certain weapons so that evil-doers are not empowered to kill even more people. The third and fourth points the authors dispute can be summed up as a single point—given that autonomous weapons are inevitable, we should be investing our time making them more ethical rather than outlawing them. To this the authors contend that autonomous weapons fundamentally change the combat landscape, and what would have been ethical before may not be ethical in this new landscape.

1.3 We Should Not Ban Killer Robots, and Heres Why

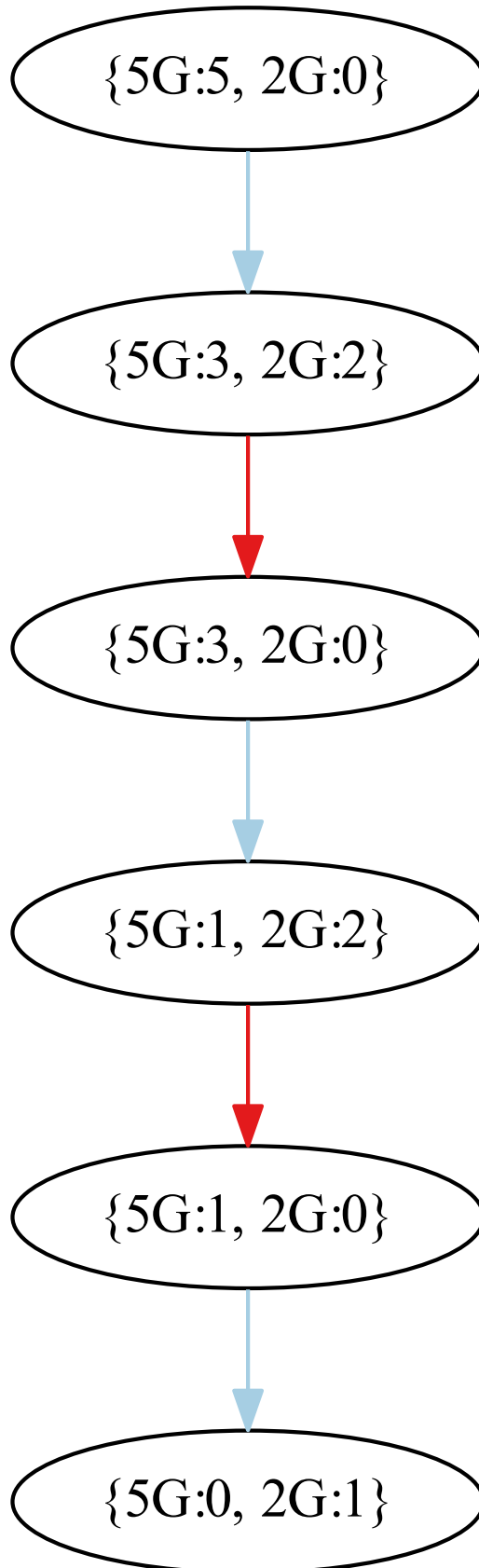
We Should Not Ban Killer Robots, and Heres Why The previous article was a response to this one. The author argues that the barrier to entry for acquiring autonomous weapons will be far too low in the very near future so a ban will not prevent people who

want them from acquiring them. From this he advocates a sort of least-harm attitude toward autonomous weapons—if you can’t prevent people from acquiring them, make it more difficult to kill en masse when they do. From there he argues that robots could be programmed to be more ethical agents than their human counterparts because they can be programmed to perfectly follow, for example, the rules of engagement. Whereas human beings are notorious for imperfectly following rules intended to ensure ethical behavior. His final point is that advocating for a ban shifts the blame for human’s unethical behavior over to machines. His main point however is that we should be having a discussion about the merits and potential risks of autonomous weapons rather than prematurely advocating for their ban.

1.4 Autonomous Weapons: an Open Letter from AI Robotics Researchers

In the article above, the author is responding to an open letter presented at the *International Joint Conference on Artificial Intelligence* in Buenos Aires, Argentina. In his article, he included the full text of the letter so here are its main points. The letter argues that the low cost barrier to entry for autonomous weapons will cause a global arms race as soon as they are available. Because of this, it is our responsibility to act now because once this arms race begins, it will only be a matter of time before autonomous weapons end up in the hands of terrorists and the like. The authors compare their advocacy for a ban on autonomous weapons as akin to the physicists who advocated for a ban on nuclear weapons when they were in development. The final point of the letter was that, as AI researchers, they see AI as having “...great potential to benefit humanity...” and that starting an arms race could potentially “...create a public backlash that curtails its future societal benefits...”

Optimal Solution



State Space for Jug Problem

