CS3105 Practical 3

170011474

2020 May

Q1 Uncertainty

(a)

(i)

When the weather is rainy $\wedge\neg$ windy or $\neg$ rainy $\wedge$ windy, I would lose money.

(ii)

1. $P(\text{rainy}) = \frac{1}{(1+2)} = \frac{1}{3}$, $P(\neg\text{rainy}) = \frac{2}{(1+2)} = \frac{2}{3}$;
2. $P(\text{windy}) = \frac{1}{(1+1)} = \frac{1}{2}$, $P(\neg\text{windy}) = \frac{1}{(1+1)} = \frac{1}{2}$;
3. $P(\neg(\text{rainy} \vee \text{windy})) = \frac{1}{(1+5)} = \frac{1}{6}$, $P((\text{rainy} \vee \text{windy})) = \frac{5}{(1+5)} = \frac{5}{6}$;
4. $P(\text{rainy} \wedge \text{windy}) = \frac{1}{(1+5)} = \frac{1}{6}$, $P(\neg(\text{rainy} \wedge \text{windy})) = \frac{5}{(1+5)} = \frac{5}{6}$;

(iii)

Kolmogorov's Axioms

1. For any event A, $0 \leq P(A) \leq 1$.
2. $P(true) = 1$, $P(false) = 0$.
3. $P(\neg a) = 1 - P(a)$.
4. $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$.

From the implied probability of the first and the second bet, we know $P(\text{rainy}) = \frac{1}{3}$ and $P(\text{windy}) = \frac{1}{2}$. $P(\text{rainy} \vee \text{windy}) = P(\text{rainy}) + P(\text{windy}) - P(\text{rainy} \wedge \text{windy}) = \frac{1}{3} + \frac{1}{2} - \frac{1}{6} = \frac{2}{3}$. However, according to bet 3, we know $P(\text{rainy} \vee \text{windy})) = 1 - \frac{1}{6} = \frac{5}{6}$. Since $\frac{2}{3} \neq \frac{5}{6}$, it obeys Kolmogorov's Axioms.

(iv)

4 bets with amount of money changed:

1. It rains vs no rain: £2 plays £4.
2. It is windy vs no wind: £3 plays £3.
3. NEITHER rainy NOR windy vs EITHER windy OR rainy: £1 plays £5.
4. BOTH rainy AND windy vs EITHER no rain OR no wind: £1 plays £5.

If I bet £1 on the first 3 bets and bet 5 on the last bet. When the weather is:

1. rainy and windy: wins $4 + 3 - 1 - 5 = 1$.
2. rainy and not windy: wins $4 - 3 - 1 + 1 = 1$.
3. not rainy and windy: wins $-2 + 3 + -1 + 1 = 1$.
4. not rainy and not windy: wins $-2 - 3 + 5 + 1 = 1$.

So the net gain would always be 1 pound.

(v)

A Dutch book is a set of odds and bets which guarantees a profit, regardless of the outcome of the gamble. It is associated with probabilities implied by the odds not being coherent, namely are being skewed.

When we simply get probabilities wrong, they might be counter-intuitive. But as long as it obey the Kolmogorov's rule, no ditch book can be made. However, if our estimates are irrational because they disagree with the Kolmogorov's rule, then there must be a dutch book where we always lose money.

(b)

Partial Observability

Observability means to what extent an external sensor can see the state of the environment to make an optimal decision.For a partial observable system, the entire state of the system is not fully visible. In a partially observable system the observer may utilise a memory system in order to add information to the observer's understanding of the system.

In this case, the problem raises when the AI of a snow plough needs to schedule a path to get around obstacles with in a minimum traveling length. Since the environment is partially observable for the AI, because it does not understand how large an obstacle is and the scene behind the current obstacle, the whole state is unknown for it. The designers need to consider partial observability and design a good algorithm to overcome it.

Noisy Sensors

Noisy sensors are sensors that have noisy data when they are collecting data from external systems. Noisy data is data with a large amount of additional meaningless information in it called noise. Sometimes data that cannot be understood or interpreted correctly is also included in noisy data. The noise can come from external sources, hardware noise, or environmental effect such as temperature.

In this case, if sensors cannot properly handle noisy data, the amount of storage space required would unnecessarily increases and the state information of external systems would be messed up. Such errors may seriously impact the answer to any query posed to the sensors. When the AI is using a camera sensor to determine whether the front area is covered by snow, a noisy sensor may return "no" due to a large amount of noisy data. Therefore, the designers need to think about how to handle noisy data and how to reduce noisy data that would occur.

Uncertainty in Action Outcomes

Different possible actions result in different outcomes. Besides, each action can have different outcomes with each outcome associating with its own probability. This probability distribution brings uncertainty in action outcomes.

In this case, AI in the snow plough would have multiple choices of actions and also have their related uncertainty in action outcomes. AI needs to make an optimal decision based on the uncertainty. Designers need to make a relatively good decision-making algorithm. Otherwise, the AI cannot fulfill its goal, reducing working efficiency.

Modelling Complexity

The modelling complexity represents how complex a model can be. It can include the number of features and properties to be considered for a decision-making process, or parameter types involved.

In this case, if a model is too simple for AI to schedule a path, then the model considers less factors. If it meets an obstacle, then it probably heads to a direction which requires a longer travelling path to get around the obstacle. If the AI wants to recognise whether or not the front area is a snow area to be cleaned, a too simple model would decrease the accuracy. A too complex model, however, also brings over-fitting problem and increases computation difficulty. So the designers of AI need to consider the modelling complexity trade-off to avoid those problems.

(c)

(i)

$P(\text{snow}) = 0.23 + 0.08 + 0.06 + 0.22 = 0.59.$

(ii)

$X = 1 - 0.23 - 0.08 - 0.06 - 0.22 - 0.23 - 0 - 0.1 = 0.08.$

(iii)

$P(\text{snow}|\text{not safe}) = \frac{P(\text{snow} \wedge \text{not safe})}{P(\text{not safe})} = \frac{0.08+0.22}{0.08+0.22+0+0.1} = \frac{3}{4} = 0.75.$

(iv)

$P(\text{not clear}|\text{snow or safe}) = \frac{0.06+0.22+0.08}{0.23+0.06+0.08+0.22+0.23+0.08} = \frac{2}{5} = 0.4.$

Q2 Attacks on AI

(a)

General AI is competent at multiple tasks similar to a human multi-task capability, while strong AI is held to have a real mind as opposed to a partial model of one and they can have consciousness and cognitive states. For example. an AI that can do a lot of things as human but cannot feel its own happiness is a general AI but not a strong AI.

(b)

Nowadays, the strong AI faces many difficulties but it is not impossible to be realised, although Seale proposed a Chinese Room Argument(CRA) to show Turing Test's flaw and hence foreshadowed a bottleneck of strong AI. The logical arguments of the CRA focuses too much on biological structures with other possibilities neglected, so they are not convincing enough. Besides, some are fallacious. Therefore, there are obstacles, but there are still many ways.

(c)

Searle introduced a Chinese room argument (CRA), which states that questions can be written in Chinese and sent into a room. An operator in that room can process Chinese questions by referring to a formal rule book and mapping symbols to feedback answers in English. Seale's logical argument is that the operator himself does not need to understand Chinese but knows how to manipulate symbols based on shapes without mental states involved. In this way, it can still pass the Turing Test but never really think intelligently. Besides, Searle argues that "brains cause minds", thought can only arise from material which is similar to our biological brain. Computer hardware only provides electrical portions without chemical and neuron parts, and hence it would never successfully realise the intelligence of human beings. This thesis strongly disagrees with the strong AI, since strong AI is AI that can have its own cognitive states as well as consciousness. Therefore, strong AI is unreachable and a bottleneck for AI is set up.

Functionalism presents that mental states are identified by what they do rather than by what they are made of. This theory is reflected by many replies and is opposite to the "biological brain" supported by Searle. The essence of his CRA is against computational versions of functionalism, particularly against functions that only focus on inputs and outputs without internal states.

Nevertheless, there are some scholars proposed replies to Searle's pessimistic attitudes towards strong AI. The most famous one is the system reply. Its logic concedes that the operator does not understand Chinese but the system understands, and the operator is part of the system, and hence understanding is still made. This argument is refuted by Searle by a "internalise system inside the operator"(i.e. remember the rule book and do symbol manipulation outside of the room). Consequently, the operator still does not understand

Chinese, so nothing in the system understands. But the assumption Searle made "if one does not X, then no part of it does X" is false since "a house is not red" does not imply "a red chair inside the house is not red". Hence, his assumption does not apply to this argument.

In addition to the system reply, the robot reply states that the reason why the operator in CRA is not intelligent is that there are no sensors, such as cameras or microphones, to facilitate a causal interaction with real worlds, so it cannot understand semantics(Piek, n.d.). It shifts from traditional AI approaches to embodiment approach, based on situatedness which means the physical world directly affects the behaviour of a robot. In this way, the connection is semantic rather than merely syntactical. Searle rejects it by saying that images the robot received are stored in binary format, without understanding made. However, the working mechanism of human memory is based on different strengths of connections between neurons via activated synapses. This mechanism is still another data encoding rather than an image itself, so no understanding is made during this process but the whole brain still understands. His logic fails again. Therefore, the robot reply still holds.

Besides, the Brain Simulator reply's logical argument is that using a program to simulate the actual sequence of firings occurring in a brain of a native Chinese speaker, so the same processing way is duplicated. In this way, Chinese is understood. Searle argues that if valves and water pipes are adopted in the same arrangements as the neurons in the Chinese speaker's brain, the program used tells how to turn on and turn off them according to inputs. There is still no understanding made. However, in his statement, the whole biological brain is intelligent, but no neuron inside actually understands Chinese as well. Similarly, the simulated sequence itself does not necessarily need to understand Chinese, as long as the whole does. As a consequence, his analogy is fallacious and this reply still holds.

Nowadays, we do encounter troubles in further developing deep learning techniques and more advanced AI. Those difficulties include that deep learning can only do specific tasks (Oster, 2019) and has fundamental limitations to be generalised. It cannot evolve to strong AI since it relies on a statically pre-trained model with parameters tuning and layers, which shares few characteristics with genuine neurons. It also heavily depends on datasets and computational resources, often affected by data cleaning. However, simply assuming intelligence can only be organic is unreasonable. An artificial brain does not need to retain all characteristics of a human brain, but only those which can help to accomplish tasks. Different models can be adopted for different units of the brain, as long as they can assemble all components into one artificial brain (Oster, 2019). In his article, recently discovered techniques including Spiking Neural Network, Hebbian Learning, etc. can be utilised.

Admittedly, CRA does raise some potential issues of development of AI, but CRA became flawed after those replies came out. In conclusion, an AI winter is unlikely to happen.

References

[1] Biermann, C., 2005. *System Reply: A Means To The End Of The Chinese Room.* [online] Citeseerx.ist.psu.edu. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.557.3201&rep=rep1&type=pdf [Accessed 16 May 2020].

[2] Copeland, J., 2000. *What Is Artificial Intelligence?.* [online] Alanturing.net. Available at: http://www.alanturing.net/turing_archive/pages/reference%20articles/what_is_AI/What%20is%20AI13.html [Accessed 16 May 2020].

[3] Cordio, M., 2008. *Searle's Chinese Room Argument And Its Replies: A Constructive Re-Warming And The Future Of Artificial Intelligence.* [online] Available at: https://cogs.sitehost.iu.edu/icogsci/vol3/Searle's%20Chinese%20Room.pdf [Accessed 15 May 2020].

[4] Oster, B., 2019. *Practical Tips For Developing An Artificial General Intelligence.* [online] AiThority. Available at: https://www.aithority.com/guest-authors/practical-tips-for-developing-an-artificial-general-intelligence/ [Accessed 16 May 2020].

[5] Piek, M., n.d. *The Chinese Room And The Robot Reply.* [online] Arno.uvt.nl. Available at: http://arno.uvt.nl/show.cgi?fid=146015 [Accessed 16 May 2020].