TDT4137 (Cognitive Architectures) Assignment Sheet 1

Danilas Miscenko

September 12, 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 Newell and Simon: Symbols and Search

Allen Newell and Herbert Simon postulate in their 1975 Turing Award lecture two hypotheses (which they developed and refined over several decades):

- the Physical Symbol System hypothesis
- the Heuristic Search hypothesis

Physical symbol system hypothesis

Describe or explain, *in your own words* (means: do not copy longer text extracts)

Task 1.1: What is a *symbol*, from the perspective of Newell and Simon?

A *symbol* is an entity that consists of some sort of pattern. An familiar example of a symbol might be a word from the English language. Each word in the language is a symbol in this context.

Task 1.2: What is a *symbol structure?*

A *symbol structure* (also called *an expression*) is just a collection of symbols that somehow physically relate to each other, like for example being next to each other. The instances of symbols in a symbol structure are also called *tokens*. If we again look at the English language as an example, a symbol structure would be a sentence, since it is a collection of words. The words in a sentence would therefore be the tokens of the symbol structure.

Task 1.3: What does the word *'physical'* mean in this context?

In this context, *physical* means that the system obeys the laws of physics, or in other words, it's possible to create such a system in the physical world.

Task 1.4: Besides symbols, what is the second, equally important component of a symbol system?

Besides symbols, the system also has a collection of *processes*.

Task 1.5: Explain what this second component does

These processes operate on expressions to produce other expressions. The collection of processes includes *creation*, *alteration*, *reproduction* and *deletion*. This basically means that the processes are used to read an expression

before executing some sort of action based on the expression. This action will then result in either a new expression, or the destruction of the expression.

Task 1.6: What is, according to Newell and Simon, the relation between a physical symbol system (PSS) and intelligent behaviour? (Hint: this relation has two main aspects)

The hypothesis stated by Newell and Simon is as follows:

```
A physical symbol system (PSS) has the necessary
and sufficient means of general intelligent action.
```

The two main aspects of the hypothesis are that a PSS has the necessary functionality to exhibit general intelligence, and that general intelligent action calls for a PSS.

'Symbol grounding'

Task 1.7: optional What is, according to your own opinion, necessary for being able to say that a regarded symbol, or a symbol structure, has a *'meaning'?*

Meaning in humans is generally derived from association of something that we wish to assign meaning to to something that already has meaning to us. An example of this would be learning a new skill (for example "carpentry") and assigning meaning to it based on what can be achieved by using the skill (creating furniture). Another human way of assigning meaning to things would be based on the feelings we have about the things we assign meaning to. An example of this would be the concept of speed. Initially, going fast might not mean anything to a person, but if they experience exhilaration, then speed will become synonymous with exhilaration. If the same person then would become injured in some way because of the speed at which they were going, the meaning might change to caution or fear.

The digression of humans assigning meaning to things due to feelings aside, I believe that a symbol structure has meaning if it has some sort of purpose. The structures purpose may be to execute a process, choose an action, a pointer in memory, a variable assignment or what ever it may be. As long as the structure does more than take up space in memory, it's not meaningless.

Symbolic vs. non-symbolic processing

Task 1.8: optional Give at least two different examples of non-symbolic processing (ofdata)

Nilsson mentions some examples in his discussion about the Physical Symbol System Hypothesis.

- Graphical models of the new-cortex
- Loosely-coupled society of simple computational ”agents”
- A “dynamic Bayes network” of symbolically-represented propositions

Nilsson, however, did not describe these systems in great detail.

Task 1.9: If an intelligent system interacts with the world, can all the necessary processing be symbolic?
Where would non-symbolic processing be useful (or necessary) in that situation?

I personally believe that the difference between a symbolic and non-symbolic processing system is purely semantic. What I mean is that a system may be designed with a non-symbolic architecture in mind (the human brain for example), but at the end of the day, the way data is processed by a computer will be symbolic because of the way the unit works.

That being said, non-symbolic system architectures might have a heuristic advantage over symbolic system architectures in some specific fields, such as language understanding or planning tasks. Therefore, a fusion of the two paradigms would certainly be beneficial, as one could compensate for the shortcommings of the other, and vice versa.

2 The Knowledge Level

According to Newell if we regard a system (e.g. a human being, an animal, or an 'intelligent machine') on the knowledge level, the system has a (dominant) law of behaviour. This law of behaviour has to do with *goals*, and a collection of possible *actions*.

Task 1.10: What does it mean if we say that the regarded system behaves in a *rational* manner?

If the system behaves rationally, then it generally means that it will select the *action* out of its available *actions* that brings it the closest to achieving its *goal*.

Task 1.11: optional Did Newell forget about perception? How can perception be integrated into this model?

The law of behaviour for a system we regard on the knowledge level states that the system will act rationally based on the knowledge it has about its environment. The medium being processed by this law is therefore the *knowledge* it has. By using the available *perception* mechanisms, the system can update its knowledge and keep track of the current state of the environment.

Task 1.12: optional According to Newell, if a system uses *all* of the knowledge that it has, it must be called perfectly intelligent. Still, such a system might be far from human intelligence. If we follow
Newell in his claim, there are (at least) two main obstacles that prevented the AI community (so far) from building machine that approach general human intelligence. Name these obstacles, and comment this.

Human intelligence deals with incredibly vast amounts of information all of the time. A perfectly intelligent system would have to use all of this knowledge to solve each problem, which would require vast amounts of storage and processing power, not to mention the time it would take to parse through all of the information. Imagine a system that has a goal of crossing the street. Were it perfectly intelligent, it would have to parse through all of the knowledge it has about everything it has ever learned about (to name a few examples) history and cooking, and language, and how to wash clothes, etc. before deciding to press the button to make the light at the junction green. It then repeats the same process again to decide to start walking when the light turns green.

This time issue also touches upon another problem, which is the idea of bounded rationality. Basically, if the rational decision takes too long to compute, it becomes irrational. The system must then either execute the now irrational action, or choose a different, less rational action. Either way, the system irrational itself becomes irrational.