Some Philosophical Problems from the standpoint of Artificial Intelligence : A Summary

Ankit Goyal (12120), Enayat Ullah (12407) and Swapnil Jain (12749)

IIT Kanpur, Kanpur, India-208016

Abstract. We have tried to compile a brief summary of the paper 'Some philosophical problems from the standpoint of artificial intelligence' by John McCarthy and Patrick J. Hayes. John McCarthy has been a pioneer in both artificial intelligence and philosophical logic. And this critically acclaimed work by him extensively explores the intersection between both the disciplines. In our work we have attempted to put forth a concise version of this 50 page paper while keeping the main ideas intact.

Keywords: Artificial Intelligence, Philosophical Logic

1 Introduction

Some of the problems of philosophy are included in Artificial Intelligence. It involves programs that perceive the environment and performs actions so as to achieve the desired goal. This involves the concepts of reasoning, knowledge, ability and causality and formalizing it. These things are also a part of philosophical logic and this is where the need of philosophy comes in Artificial Intelligence.

2 Philosophical Questions

2.1 Why Artificial Intelligence needs Philosophical Logic?

Artificial Intelligence involves developing systems endowed with the intellectual processes characteristic of humans such as ability to reason and other things. Still inspite of the development, it lacks behind in creating programs with the same kind of flexibility as that of humans.

An entity is said as intelligent if it has the capability to model a world, understand and analyse it and based on the facts and other information achieve the desired goal. According to this, Intelligence comprises of two parts:

1. Epistemological- It entails to the representation where solution follows from the facts given
2. Heuristic- It solves the problems by analysing the situation and taking decision based on the information.

The paper mainly deals with the epistemological part of Artificial Intelligence. However there are certain problems which come while constructing it like how to interlink the physical world with other entities that are provided like knowledge, the desired goal etc. Many of these issues also form a part of the traditional topics in philosophy especially in the field of metaphysics, epistemology and philosophical logic.

2.2 Reasoning programs and Missouri program

Reasoning program is a kind of intelligent program which interacts with the world through the input output devices and has a representation for information that it gathers in a variety of ways. However mostly one representation plays a dominant role and that is decided by taking into consideration various factors. Missouri program on the other hand deals only with the epistemological part of the intelligence and tries to execute actions so as to achieve the desired results.

2.3 Representations of the world

The representation of the world implies deciding the structure of the world. Three main representations have been discussed:

1. Metaphysically adequate: If world has a form that doesnt contradict reality.
2. Epistemologically adequate: If we can express the actual facts
3. Heuristically adequate: If the reasoning process can be expressed in language

2.4 Automaton Representation and Notion of Can

Automaton Representation is having systems which interact which each other. It consists of finite automata which interact with each other and these constitute a automaton representation. However certain difficulties arise due to the fact that it is epistemologically inadequate as the initial states are not well known. However this representation is used for the concepts of can, cause, counterfactual statements and statement of believe.

3 Formalism

In this section, the construction of an epistemologically adequate system is discussed, wherein formal notions are introduced using informal natural-language descriptions and examples of their use to describe situations and the possibilities for action they present are given.
Some definitions are required to address the task above:

3.1 Situation

A situation s is a complete state of the universe at a given time. We use Sit to denote a set of all situations. Since it is not viable to completely describe the universe at a given time, we restrict ourselves to conveying facts about the situation. Further, we also consider hypothetical situations which deal with conditionals such as 'What happens if X does Y'. It is further assumed that the laws of motion determine, given a situation, all future situations.

3.2 Fluent

A fluent is function whose domain is the space Sit of all situations. The range encompasses two possibilities:
1. Propositional Fluent: The range is $True, False$.
2. Situational Fluent: The range is the space Sit.
For example: $raining(x)(s)$ or $raining(x, s)$ is a propositional fluent whose value is $True$ if it is raining at the place x in the situation s.
A rather complex representation would involve a situation s that person p is in place x and that it is raining in place x.

$$at(p, x, s) \wedge raining(x, s)$$

Moreover, if the concept of λ-abstraction is employed, the above would translate into:

$$[\lambda s'.at(p, x, s') \wedge raining(x, s')](s)$$

3.3 Causality

The concept of causality is incorporated using the fluent $F(\pi)$, where π is a propositional fluent. $F(\pi, s)$ asserts that the situation s will be followed by a situation (after an unspecified time) wherein the fluent π is satisfied.
For example: The situation "If a person is out in the rain, he will get wet"; is represented as:

$$\forall x.\forall p.\forall s.raining(x, s)at(p, x, s)outside(p, s) \rightarrow F(\lambda s'.wet(p, s'))$$

Besides F, there are three other operators which are useful in implementing the notion of causality:

- $G(\pi, s)$. For all situations s' in the future of s, $\pi(s')$ holds.
- $P(\pi, s)$. For some situations s' in the past of s, $\pi(s')$ holds.
- $H(\pi, s)$. For all situations s' in the past of s, $\pi(s')$ holds.

It is also useful to define a situational fluent $next(\pi)$, which outputs a situation s' in the future of s, in which $\pi(s')$ holds. In case, there is no such situation, $next(\pi)$ is considered undefined.Following is an example which encapsulates the

notion of next:
"By the time John gets home, Henry will be home too".

$$at(Henry, home(Henry), next(at(John, home(John)), s))$$

But the computation of next is rather intensive, since it requires a rich description of the situation, so we rather have a fluent time(s), wherein the values of fluent applied to next is computed. The above statement thus is represented as:

$$at(Henry, home(Henry), time(next(at(John, home(John)), s)))$$

3.4 Actions

Before actions, we must consider the situational fluent $result(p, \sigma, s)$, where p is a person, σ is an action and s is a situation. $result(p, \sigma, s)$ outputs a situation that occurs when p carries out σ, starting in the situation s. In case the action does not terminate, $result(p, \sigma, s)$ is considered undefined.
An example to illustrate the above is: If in a situation s a person p has a key k that fits the safe sf, then in the situation resulting from his performing the action $opens(sf, k)$, that is, opening the safe sf with the key k, the safe is open.

$$has(p, k, s) \wedge fits(k, sf) \wedge at(p, sf, s) \rightarrow open(sf, result(p, opens(sf, k), s))$$

3.5 Strategy

A strategy is a finite sequence of actions performed in a specific order to achieve an end. A simple notation to express a strategy is using ALGOL statements or procedure calls. For example: the sequence of actions: move the box under the bananas, climb onto the box, and reach for the bananas sequence is expressed as:

```
begin
$move(Box, UnderBananas)$;
$climb(Box)$;
$reachFor(Bananas)$
end;
```

3.6 Knowlege and Ability

Knowledge plays an important role towards one's ability to achieve goals. Let us revisit the example "If in a situation s a person p has a key k that fits the safe sf, then in the situation resulting from his performing the action $opens(sf, k)$, that is, opening the safe sf with the key k, the safe is open".

$$has(p, k, s) \wedge fits(k, sf) \wedge at(p, sf, s) \rightarrow open(sf, result(p, opens(sf, k), s))$$

Now, if in place of a key safe, we have a combination safe with a combination c, and we use the predicates $fits2$ and $open2$ to express the distinction between a key fitting a safe and a combination fitting it.

$$fits2(c, sf) \wedge at(p, sf, s) \rightarrow open(sf, result(p, opens2(sf, c), s))$$

Note that the predicate $has2(p, c, s)$ is left out. It is primarily done because if we manipulate a safe in accordance with its combination it will open; there is no need to have anything. Also, it is not very clear what $has2(p, c, s)$ mathematically means. Thus, a direct parallel between the rules for opening a safe with a key and opening it with a combination seems impossible. We first modify the expression as follows:

$$at(p, sf, s) \wedge csafe(sf) \rightarrow open(sf, result(p, opens2(sf, combination(sf)), s))$$

Here $csafe(sf)$ asserts that sf is a combination safe and $combination(sf)$ denotes the combination of sf. Next, we move on to expressing the fact that one has to know the combination of a safe in order to open it. The first approach is to regard the action $opens2(sf, combination(sf))$ as infeasible because p might not know the combination. Therefore, we introduce a new function $idea - of - combination(p, sf, s)$ which stands for person p's idea of the combination of sf in situation s. The assertion that p knows the combination can be expressed as:

$$idea - of - combination(p, sf, s) = combination(sf)$$

A further consequence of our approach is that feasibility of a strategy is a referentially opaque concept since a strategy containing $idea - of - combination(p, sf, s)$ is regarded as feasible while one containing $combiantion(sf)$ is not, even though these quantities may be equal in a particular case.

4 Problems and Remarks on Formalism:

The formalism presented above, though efficient to represent most of the situations, is far from epistemological adequacy. Following is list of diverse problems which it fails to cater to:

4.1 The approximate character of $result(p, \sigma, s)$

Consider the situation in which if someone is asked, 'How would you feel tonight if you challenged him to a duel tomorrow morning and he accepted?' he might well reply, 'I cant imagine the mental state in which I would do it; if the words inexplicably popped out of my mouth as though my voice were under someone elses control that would be one thing; if you gave me a long-lasting belligerence drug that would be another.'

From the above, we see that $result(p, \sigma, s)$ should not be regarded as being defined in the world itself, but only in certain representations of the world. We regard this as a blemish on the smoothness of interpretation of the formalism, which may also lead to difficulties in the formal development.

4.2 Possible Meanings of 'can' for a Computer Program

A computer program can address to various notions of various notions of $can(Program, \pi)$. It can readily be given much more powerful means of introspection than a person has, for we may make it inspect the whole of its memory including program and data to answer certain introspective questions, and it can even simulate (slowly) what it would do with given initial data.

4.3 The Frame Problem

Suppose we have a number of actions to be performed in sequence, then we would have quite a number of conditions to write down that certain actions do not change the values of certain fluents. In fact with n actions and m fluents we might have to write down mn such conditions.
The notion of frames(State vector in McCarthy 1962) resolves this issue. A number of fluents are declared as attached to the frame and the effect of an action is described by telling which fluents are changed, all others being presumed unchanged.
Formally, consider a strategy in which in which p performs the action of going from x to y. It can be expressed by writing either $go(x, y)$ or $s := result(p, go(x, y))$. However, if we write $location(p) := tryfor(y, x)$, the fact that other variables are unchanged by this action follows from the general properties of assignment statements. The point of using tryfor here is that such a program may not be possible to execute, since p may be unable to go to y. This case may be covered in a more complex assignment by agreeing that when p is barred from y, as $tryfor(y, x) = x$.

4.4 Probabilities

It is suggestive that formalism should take uncertainty into account by attaching probabilities to its sentences. However, the information necessary to assign numerical probabilities is not ordinarily available. Therefore, a formalism that required numerical probabilities would be epistemologically inadequate.

4.5 Parallel Processing

Instead of describing strategies using ALGOL-like programs, we must take into account the fact that many processes are going on simultaneously and that the single-activity-at-a-time ALGOL-like programs will have to be replaced by programs in which processes take place in parallel, in order to get an epistemologically adequate description. Simulation languages is a domain which allows for such strategies, however, they are rather restricted in the kinds of processes they allow to take place in parallel and in the types of interaction allowed.

5 Discussion of Literature

This section is devoted to some of the simpler schemes proposed.

- L. Fogel (1966) proposal to alter state transitions is limited to automata with small number of transitions and thus not adequate. Representation of behavior in terms of states may be metaphysically adequate but it lacks the epistemological adequacy as what is learned from experience cannot be expressed in terms of fixed states.
- Galanter (1956), Pivar and Finkelstein (1964) view of intelligence as ability to predict future from past events is metaphysically adequate but epistemologically inadequate. Even humans cant predict the exact sequence of events from the past experience.
- Friedberg (1958,1959) approach of representing behavior by computer program and evolving it with random mutations is epistemologically inadequate as desired changes in behavior often cannot be represented by small changes in machine program.
- Newell and Simons General Problem Solver attempts to solve problems by representing them in another form. But representation of many problems was awkward enough for the GPS to solve it.
- Newell and Ernst (1965) view that the class of problems the problem solver can solve depends on its ability to represent the external to its internal. The division of problem solver into program that converts the external into internal and the program that solves the internal representation is very similar to the division into heuristic and epistemological parts of an artificial intelligence.

6 Epistemological representation

Some powerful tool for the epistemological representation of an artificial intelligence problems have been briefly described in the preceding sections.

6.1 Modal Logic

This system of logic stems its origin from the works of C. L. Lewis. The main idea was to distinguish necessary truths from the contingent ones. He introduced the modal operators ($\Box$ necessity and $\diamond$ possibility). Following axioms are added to propositional calculus to get M logic system.

1. $\Box p \to p$
2. $\Box(p \to q) \to (\Box p \to \Box q)$

M is weak modal and can be strengthened by adding several axioms leading to $S4$, $S5$. The axioms can also be weakened if required. Modal logic turn out to be powerful tool to analyze logic of various operators such as knowledge, tense, belief. Initial modal logic suffered from problems like failure to satisfy Lebniz's law of identity. The possible word semantics of modal developed chiefly by Kripke provides a more satisfactory method for interpreting modal sentences. The main idea behind the Kripke semantics is that modal calculi describes several possible worlds at once rather than just one. And thus statements are assigned a spectrum of truth values, one for each possible world. This takes care of certain problems still certain difficulties prevail.

6.2 Logic of Knowledge

It was first investigated as a modal logic by Hintikka in his book Knowledge and belief (1962). He introduced the modal operator K_a (read 'a knows that'), and its dual P_a, defined as $\neg K_a \neg$. The semantics is obtained by the analogous reading of K_a as: 'it is true in all possible worlds compatible with a's knowledge that'. This analysis of knowledge has been criticized in various ways, to which Hintikka has replied in his important papers.

6.3 Tense Logic

This is one of the most active areas of philosophical logic. Past, Present and Future by Prior provides a thorough account of this field. He discusses the four logic operators F, G, P and H. These are regarded as modal operators for the time ordering operations. Tense logic with time has reached high level of technical sophistication.

6.4 Logics and Theories of Action

Von Wright in his book Norm and Action (1963) presented the most developed theory in this area. He builds his logic on a rather unusual tense-logic of his own. The basis is a binary modal connective T, so that pTq, where p and q are propositions, means p,then q. Von Wright (1967) later altered and extended his formalism so as to answer some criticism and also has provided a sort of semantic theory based on the notion of a life-tree.

7 Conclusion and Remarks

The paper 'Some philosophical problems from the standpoint of artificial intelligence' is one earliest works in the fields in artificial intelligence and one can easily trace how ideas of philosophical logic has influenced research in the field of AI. Although AI research today is more concentrated towards developing rigorous mathematical models, logic has and will always remain an important tool to understand and model reasoning in intelligent beings. In that sense this paper by McCarthy is of utmost importance for the AI community. We have tried to present this paper in a concise and crisp form while giving due stress to the main ideas and concepts.

8 References

1. McCarthy, John, and Patrick Hayes. Some philosophical problems from the standpoint of artificial intelligence. USA: Stanford University, 1968.
2. Hintikka, J. (1967c). Individuals, possible worlds and epistemic logic. Nous, 1, 33-62.
3. Kripke, S. (1963a). Semantical considerations on modal logic. Acta Philosophica Fennica, 16, 83-94.
4. Lewis, C.I. (1918). A survey of symbolic logic. Berkeley: University of California Press.
5. Fogel, L.J., Owens, A.J. and Walsh, M.J. (1966). Artificial Intelligence through simulated evolution. New York: John Wiley.
6. McCarthy, J. (1962). Towards a mathematical science of computation. Proc. IFIP Congress 62. Amsterdam: North-Holland Press
7. von Wright, C.H. (1963). Norm and action: a logical enquiry. London: Routledge
8. von Wright, C.H. (1967). The Logic of Action - a sketch. The logic of decision and action (ed. Rescher, N.). Pittsburgh:University of Pittsburgh Press
9. Newell, A. and Simon, H.A. (1961). GPS - a program that simulates human problem-solving. Proceedings of a conference in learning automata. Munich:Oldenbourgh
10. Newell, A. and Ernst, C. (1965). The search for generality. Proc. IFIP Congress 65.
11. Friedberg, R.M., Dunham, B., and North, J.H. (1959). A learning machine, part II. IBM J. Res. Dev., 3, 282-7.
12. Galanter, E. and Gerstenhaber, M. (1956). On thought: the extrinsic theory. Psychological Review, 63, 218-27.

$\approx$ 3000 words