

Intelligence means to be rewarded

V. Beneš

University of West Bohemia, Plzeň, Czech Republic

keywords: intelligence, behavior, action selection, situation, umwelt

I. Introduction

In this article we will be dealing with the notion of intelligence. Intelligence is explored and a definition of what it means to be intelligent is presented. We study how could be developed an effective behavior in an environment. There are several abstract ‘mechanisms’ which are believed to help in developing effective behavior and which could be observed in human activity, too. These mechanisms are discussed along with their possible realization in an intelligent agents’ thinking machinery.

Additionally we will try to address some of the mechanisms as the bridge between embodied systems and symbolic processing. Specifically we deal with the problem of perception (defining and recognizing situations) and the problem of meaning.

In the next section we will introduce definition of intelligence and discuss it briefly. Section III is about some mechanisms which seems to be involved in creating behavior, Section IV describes an example of virtual environment and issues bonded with agent activity in it. Section V speaks about dividing world into situations – about perception and about giving meaning to things. Section VI draws some conclusions.

II. Intelligence

Present definitions of intelligence use words like “understand”, “meaning”, “knowledge”, “goals”, “comprehending”, “reasoning”. However, these words can not be easily explained in technical terms.

We define intelligence as follows:

intelligence is the ability to develop and perform behavior leading to high reward

The terms used in this definition can be translated into features of technical realization of an intelligent system (agent). We will speak about it later in this chapter. There are some interesting consequences of this definition. Among others – in simple worlds normally non-intelligent things/systems will be considered intelligent.

Embodiment

Embodiment is a mutual influence between an agent and the world around. The body has sensors, performs actions and it receives rewards. It is believed ([5],[6]) that body is needed for intelligent mind. Thinking machinery and intelligence is developing from reward-guided interplay between actions and sensing. Reward is the evaluation of the agents’ performance in the environment (world). In nature you are rewarded if you survive, if your genes/memes survive. In artificial worlds the reward for agents can depend on arbitrary features of the world. Reward system ‘describes’ goals for the agent.

By the term ‘behavior’ we mean simply the sequence of actions, produced by (the body of) the agent.

III. Machinery for developing behavior

Present artificial intelligence technologies lack ability to explore and understand the world. AI systems are specialized and their performance horribly drops if they encounter unforeseen conditions. Reason for this is that the human designers incorporate their knowledge into system without ensuring that the system is able to derive new knowledge after significant changes in the environment.

We try to solve this by letting the system to explore and understand the whole world without assistance. It means that the system will gather knowledge and useful skills like: how to find best behavior in a simple case; how to make simple case from a complex one; how to behave after something went wrong; etc.

Here we list various mechanisms which are believed to help with creating reasonable behavior:

- mechanism 1: perform random actions;
- mechanism 2: perform better actions more often;
- mechanism 3: collapse more actions into one (macro-action);
- mechanism 4: determine the ‘usual’ effect of actions;
- mechanism 5: define situations (find regions in state space where certain action performs better or where reward is high);
- mechanism 6: discover goals;
- mechanism 7: gather rules and make model of the environment;
- mechanism 8: planning;
- mechanism 9: derive useful strategies of behavior (depending on situations);
- mechanism 10: find agents’ inner actions (inner processes of an agent are considered new environment where another copy of agent could be planted – creating positive feedback loop control hierarchy);
- mechanism 11: correction (loop detection, model inconsistency detection);
- mechanism 12: sandbox, fishpool – finding simpler cases;
- mechanism 13: extend reward function from outside to more concrete inside reward function.

More close view on low level mechanisms 1 – 7 follows:

Mechanism 1: perform random actions

If you have no information about the world around – best what you can do is to perform random actions. You keep track of the sequence of actions you performed and the rewards you gained.

Mechanism 2: perform better actions more often

After collecting some statistical material, you can find out which actions and combinations of actions led to high reward. These better actions you should perform more often. It is essential to perform worse actions too to ensure that these bad actions can not be actually good, leading to high rewards in (far) future.

Mechanism 3: collapse more actions into one (macro)action

Macroactions are the way how to avoid (partially) combinatorial explosion of number of action combinations. Short sequences of actions that were especially useful (leads to high reward, reusable) are marked as a new action and used instead of sequences of (primitive) actions.

Mechanism 4: determine the ‘usual’ effect of actions

Your actions influence the world. You need to know what is the effect of your actions – how your actions change the world state-space. This means you need to compare the history of performed actions with collected data from your sensors. Why do we talk about ‘usual’ effect? Generally, the effect of an action can be alternated by extensive amount of unknown conditions. We are working

with ‘usual’ effect – effect observed in majority of all cases of performing certain action. After determining the ‘usual’ effect you can detect ‘unusual’ effects of an action and you can try to figure out what caused this nonstandard development.

Mechanism 5: define situations

Situation is state space partition to 2 or more regions. To improve obtained reward – you should try to distinguish situations in the world and try to find best behavior in each situation-region. For details how to recognize a situation see Section V.

Mechanism 6: discover goals

You search for regions in world state-space with highest rewards.

Mechanism 7: gather rules and make model of the environment

There you describe symbolically how to transform world from one situation-region into another using actions. You search for rules that describe distinct situation-regions (to avoid the need to distinguish your situation-region by employing computationally expensive machinery used in mechanism 5).

IV. Example of virtual environment

In this chapter we will present simple computer generated world. We show how agent uses behavior-developing mechanisms in this environment. Using mechanisms described in previous section the world around is explored, key aspects and entities are identified. Knowledge is gathered and used to refine agents’ behavior to obtain higher rewards.

Examined world is an environment where the agents can move in a restricted rectangular two dimensional space (Fig. 1). Agents’ body is represented as a point in this space. Additionally the whole world is divided into more rectangular regions used to study exploratory behavior of the agent.

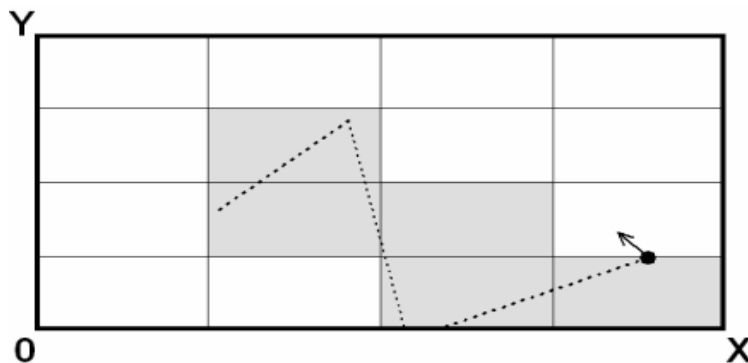


Fig. 1: Example of 2D rectangular virtual world occupied by an agent.

Table (Tab. 1) describes the key properties of the agents’ body – actions, sensors and reward function. The agent is able to perform three actions: MOVE (move 1 unit ahead), RCD (random change direction), NULL (void action). The reward system could be set entirely arbitrarily. So if we would like to see exploratory behavior of the agent – we reward agent for every newly visited region. Note that the agent is very limited in gathering information about the world. RCD action results in random change of direction the agent is heading. Because there is no sensor for direction it is difficult for the agent even to find out that the world has rectangular shape. We designed actions of the agent in this way to be able to study the lowest level of behavior in very simple setup.

AGENTS' BODY		
ACTIONS	MOVE RCD NULL	move 1 unit ahead random change direction do nothing
SENSORS	POSX POSY	X axis position Y axis position
REWARD FUNCTION	+ 1 for every newly visited region - 0.001 for performing any action	

Tab. 1: Agents' body.

Agent starts from scratch in our new environment. There is no clue what to do, so the agent randomly choose one from actions MOVE, RCD and NULL and performs it (mechanism 1). Information about used actions, obtained rewards and sensor values are stored for further statistics.

After a while there could be seen from statistics how the performed actions influenced gained reward. Agent finds out that only MOVE action leads to immediate reward – as a consequence action MOVE is performed more often (mechanism 2).

Agent begins to move in straight line – this leads to drop in performance since the agent ends at the border and positive reward can not be gained without turning. Action MOVE is getting gradually worse – until the agent performs RCD. Then next actions MOVE can gather additional positive reward¹. In this point the agent could examine statistics of performed actions and reward statistics and create some macroactions (mechanism 3).

Macroaction M1 = (10xMOVE) could prevent repeating of MOVE action and simplify behavior of the agent.

The time has come to use collected sensor information. Till now data from sensors were just numbers – now the agent need to give them meaning, the task is here to make a sensory state space partition in some useful way. For each (macro-) action there is determined its' effect on various features of the world. We examine how this action influences this feature (set of features). There can be shown that action RCD has no influence on agents' position (POSX, POSY) but action MOVE has influence on POSX, POSY (mechanism 4). More detailed look will be presented in Section V of this document.

Agent recognizes a situation with state-space regions which can be (for humans) labeled as “border” and “inside”. It is obvious that different behavior will suit for each of the regions. The agent proceeds with refining the behavior for each region using mechanisms 1 – 5. For inside-region there will the best behavior look like something like (8xMOVE, RCD), for border-region there is considerable need for turning, so the behavior here will be something like (RCD, 2xMOVE). Defining of situation and differencing behavior for each case leads to higher reward than the previous uniform behavior.

V. Perception and meaning

When an agent appears in completely new unknown world, no previous knowledge can be used – at least that low level knowledge that describes directly² situations in previously occupied environment (if any).

¹ As we see action RCD is essential although it brings no direct reward. It seems to be necessary to distribute fraction of reward to the actions performed in the past. This makes the problem of developing behavior more complex. The agent should determine how far to the future lasts effect of each action.

² We have a suspicion that there could be used some knowledge that is common for many different worlds. This so-called meta-knowledge would be probably describing some abstract high-level concepts (eg. when was something observed many times in the past, it would appear in the future too, etc.).

The agent has to employ its behavior-creating mechanisms and learn everything about the environment from scratch from interactions with it.

Here we introduce a notion of a *situation*. A region in a world state space that is recognizable for an agent.

Our deliberation starts at the idea that the agent is not able to distinguish between states of the world that seems the same. If the difference between two states have no influence on the agent (on its' body) or what is equally important – if the agent is not aware of that influence, there is no way to distinguish these two states. Since the agent has absolutely no idea about what influences what in unknown world, it cannot perceive different situations – whole world is seen as monotonous.

How to make a world state space partition? This is the same question as: How to define a new situation? New situation is environmental state space partition to regions, where the regions are notable because:

- **fitness function value here is significantly above/below normal** (eg. what the environment was like when I died; what the world was like when happen to me that I accomplished my goal; what regions in worlds' state space are plausible for me; etc.);
- **certain action (if performed here) has unusual/usual/no/high-rewarded/low-rewarded effect.**

Now we return to our virtual world. The agent examines the effect of actions – their influence on the features of the world. Dependence between using action MOVE and differences in position is found. Fig. 2 shows examples of use of MOVE action (left) and the effect of this action on the POSX and POSY agents' sensors. Points on the circle are the cases when agent didn't hit the wall when performing MOVE.

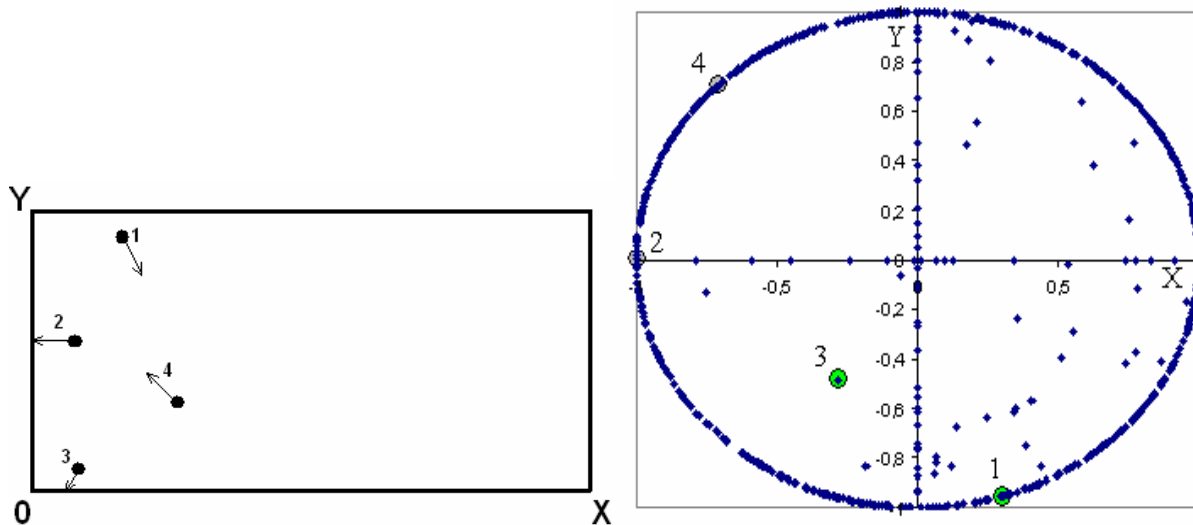


Fig. 2: left – 4 examples of effect of action MOVE; right – 1500 effects of action MOVE (difference in position – X and Y axis), 4 examples from left picture are highlighted.

In this point the question is: what is the usual effect of action MOVE? This can be answered by performing frequency based clustering ([4]) on the data (Fig. 2 – right). Fig. 3 – left shows two resulting clusters. Usual effect of action MOVE is the set of points on the circle (gray), unusual effect are all other points (green). Now we can split state space of the world into first region where MOVE has usual effect (inside) and to second region where MOVE has unusual effect (border) (Fig. 3 – right)³.

³ Note that some places with usual effect of MOVE lays in “unusual effect region” – this is caused due to the (hidden) dependence on direction the agent is heading. This is no drawback – we study unusual effects and they never occur in “usual effect region”.

As we have seen in Section IV recognizing inside/border situation allow development of more sophisticated behavior that leads to higher reward.

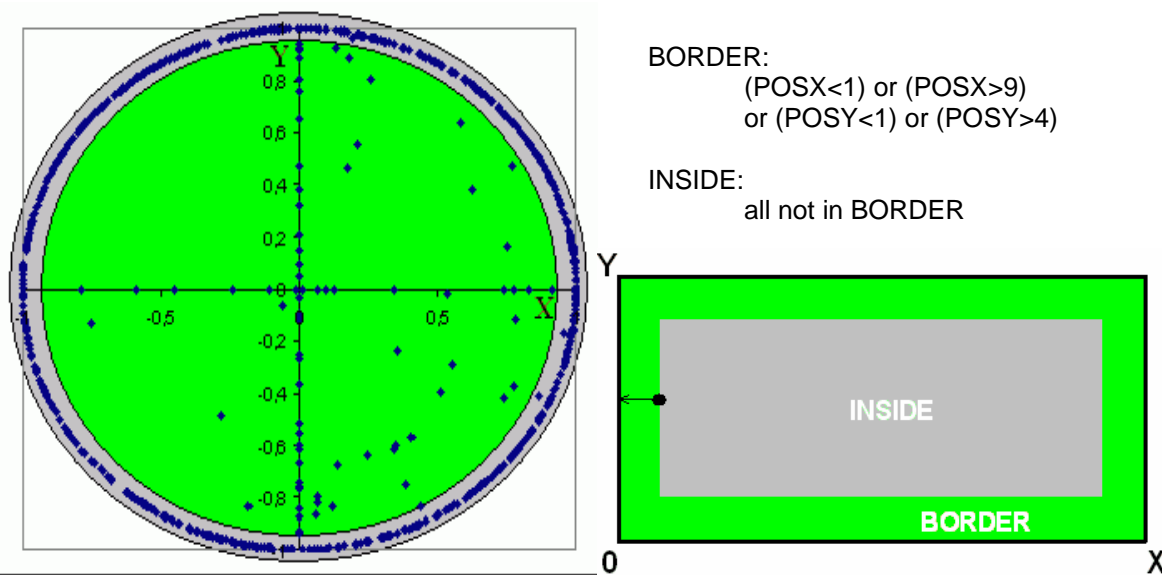


Fig. 3: recognized situation; left – effect of action MOVE after frequency clustering (gray – usual effect, green – unusual effect); right side down – inside/border situation state space partition; right side up – derived rules.

Frequency clustering is computationally expensive process. The agent needs more straightforward way how to detect actual situation-region (inside/border). Using sensor data the border (unusual MOVE effect world state space region) can be described by a set of logic rules (Fig. 3 – right).

VI. Conclusion

An agent should be considered intelligent if it manages to arrive at high gain in previously completely unknown environment. To accomplish this, the agent is likely to incorporate in its mind machinery some of the behavior creating mechanisms we discussed. We believe that described mechanisms are building blocks of intelligence. These mechanisms make possible to truly explore and understand the environment and to develop and produce high rewarded behavior.

Described mechanism of recognizing situations may bring some new insights into the symbol grounding problem.

References

- [1] McGovern, A., and Sutton, Richard S.; Macro-actions in reinforcement learning: An empirical analysis. Tech. Rep. 98-70, University of Massachusetts, Amherst, 1998.
- [2] Precup, D.; Temporal abstraction in reinforcement learning. Doctoral dissertation, University of Massachusetts Amherst, 2000.
- [3] Minsky, M.; Society of Mind. Simon and Schuster, New York, 1986.
- [4] Eldershaw, C., and Hegland, M.; Cluster analysis using triangulation. Computational Techniques and Applications: CTAC97, pages 201-208. World Scientific, Singapore, 1997.
- [5] Damasio A. R.; Descartes' Error. Emotion, Reason and the Human Brain. G. P. Putnam's Sons, New York, 1994.
- [6] Brooks, R. A.; Intelligence without reason. Proceedings of the 12th International Joint Conference on AI (IJCAI-91) (Sydney, Australia) (John Myopoulos and Ray Reiter, eds.), Morgan Kaufmann publishers Inc.: San Mateo, CA, USA, 1991, pp. 569-595.