INCORRECT MOVES AND TESTABLE STATES

Dimiter Dobrev

Abstract: *How do we describe the invisible? Let's take a sequence: input, output, input, output ... Behind this sequence stands a world and the sequence of its internal states. We do not see the internal state of the world, but only a part of it. To describe that part which is invisible, we will use the concept of "incorrect move" and its generalization "testable state". Thus, we will reduce the problem of partial observability to the problem of full observability.*

Keywords: *Artificial Intelligence, Machine Learning, Reinforcement Learning.*

ITHEA Keywords: *I. Computing Methodologies: I.2 Artificial Intelligence: I.2.6 Learning*

Introduction

Our first task in the field of Reinforcement learning is to describe the current state of the world (or the current position of the game). When we see everything (full observability) this question does not arise because in this case the input fully describes the state of the world. More interesting is the case when we see only a part of the world (partial observability). Then we will add more coordinates to the input vector, and thus we get a new vector which already fully describes the state of the world.

To add these new coordinates we will first introduce the concept of "incorrect move". The idea that not all moves are correct is not new. Other authors suggest, for example, that you cannot spend more money than you have. That is, they assume that the output is limited by a parameter that changes over time. What's new in this article is that we will use this parameter to give further description of the state of the world.

If at a specific moment we know what we see and what moves are correct at that specific moment, we know a lot about the current state of the world, yet we do not know everything. When we generalize the concept of "incorrect move", we will derive the new concept of "testable state". If we add to the input vector the values of all testable states, we will get an infinite-dimensional vector that fully describes the state of the world. That is, for every two distinct states there will be a testable state whose value in these two states is different.

Incorrect moves and testable states are something that actually exists just like the input we get on each step. However, unlike the input, the value of testable states is not ready to derive. For example, is the door locked? To check this we need to push the handle and see whether the door will open, but we can

do it only if we stand by the door. In other words, the check requires extra effort and it is not always possible (there are moments in which we can check and moments in which we can't). The locked door can be considered as an example of both an incorrect move and of a testable state.

To describe the incorrect moves and testable states we will search for a theory that gives us this description. Of course, we may have many theories for a given testable state all of which will compete with each other over time.

What will the theory constitute? Statistics shows us that in specific situations a given testable state is true. Specific situations means a situation in which a conjunction is true. This conjunction may be associated only with the past, but may also be associated with the past and the future. For example, let's say we've checked and we've seen that a door is locked, but is it the door that we are interested in? We may decide that it is precisely this door on the basis of what we have seen before checking, or maybe after checking, a posteriori, we've seen something which tells us that it is exactly this door.

Another generalization of "specific situations" will be to allow dependencies with memory (except dependencies without memory). A dependency without memory is the situation in which specific events occur in the course of several consecutive steps (i.e. this is an ordinary conjunction). A dependency with memory is the situation in which specific events occur at specific moments (not consecutive), and in the periods between those moments specific events do not occur.

The theory may be a set of such implications and this set can be generalized as a finite state machine. Let's take an example where we are moving in a castle in which some of the doors are locked. We can have a rule of the following type: "*If you see a white sofa and you turn right, then the door will be locked.*" If we represent the map of the castle as a finite state machine, we will see that if after the white sofa we turn right, we are at the door *X* and that this door is locked. If we know the finite state machine, we can easily derive the corresponding rules. Unfortunately, we need the opposite – to derive a finite state machine from the rules, and that's a much more difficult task.

Let us now imagine a castle the doors of which are not permanently locked or unlocked but change following specific rules. Then our theory will suggest some sustainability of testable states. For example, if a door has been locked at a specific moment and shortly thereafter we check again, we assume that it will be locked again, especially if during this time no specific events have occurred (for example, to hear a click of door locks).

The next upgrade of the theory would be to assume that there is some kind of creature inside the castle, which unlocks and locks the doors in its sole discretion. Then we can predict whether a door is locked or unlocked predicting the behavior of that creature.

Once we've created one or several theories that predict a testable state, we will use these theories and gather statistics that will help us predict the future and to create new theories of other testable states.

For example, in our case, if we've noticed that behind the locked door there is a princess, and a tiger behind the unlocked door, then based on the theory that tells us which door is locked, we can make a theory that tells us where the princesses are.

Definitions

Let's take a sequence of *output, input, output, input, … ,* and the goal is to understand this sequence.

Of course, this sequence is not accidental. We can assume that we are playing a game and that's the sequence:

move, observation, move, observation ...

And our goal is to understand the rules of the game and what is the current position on the game board.

We might assume that we are in a world and then the sequence would be:

action, view, action, view ...

In this case, our goal is to understand the rules of this world and what is the current state of the world.

The description of the world is given by the functions *World* and *View*, and the following applies:

$s_{i+1}=World(s_i , a_{i+1})$

$v_i=View(s_i)$

Here, actions and observations (a_i and v_i) are vectors from scalars with dimensions n and m, respectively.

Our goal is to present the internal state (s_i) also as a vector of scalars, in which the first m coordinates will be exactly v_i. In other words, the function *View* will be the projective function that returns the first m coordinates of s_i.

We will call "states" to all coordinates of s_i. We will call the first m ones "visible states". Other coordinates will be called "testable states".

The coordinates of the input and output will be called "signals". These are functions of time that return a scalar. To the input and output signals we will add other signals as well, like the testable states; more precisely – the value of the testable state according to the relevant theory, because we will not know the exact value of these states, and will approximate them with some theories. For each finite state machine we will add a signal whose value will be equal to the state in which the finite state machine is situated at the moment t. Of course, if the machine is nondeterministic, the value of this signal may not be exact but approximate.

We will call the Boolean signals "events ". When the Boolean signal is truth, we will say that the event is happening.

Example

To make things clear, let's take an example. Let's imagine we are playing chess with an imaginary opponent without seeing the board (we are playing blind). We will not specify whether the imaginary opponent is human or a computer program.

Our move will be the following 4-dimensional vector:

X_1, Y_1, X_2, Y_2

The meaning is that we are moving the piece from the (X_1, Y_1) coordinates to the (X_2, Y_2) coordinates.

What we will see after each move is a 5-dimensional vector:

A_1, B_1, A_2, B_2, R

The first four coordinates of the input show us the counter-move of the imaginary opponent, and R shows us our immediate reward.

All scalars have values from 1 to 8, except R, whose value is in the *{-1, 0, 1, nothing}* set. The meaning of these values is as follows: *{loss, draw, win, the game is not over}*.

Incorrect move

Shall we allow the existence of incorrect moves?

Our first suggestion is to choose a world in which all moves are correct. In the example we took, we cannot move the bishop as a knight, so it is natural to assume that some of our moves are incorrect or impossible.

Our second suggestion is to have incorrect moves and when we play such a move to assume that the world penalizes us with a loss. So we will very quickly learn to avoid incorrect moves, but here we are denied the opportunity to study the world by checking which move is correct (like touching the walls in the darkness).

Our third suggestion is: When an incorrect move is made the state of the world to remain the same, and instead of "loss" the immediate reward to be "incorrect move" and all other coordinates at the input to return "nothing". This option is better, but thus we would unnecessarily prolong the history. (We will call "history" the following sequence: $a_1, v_1, \ldots, a_{t-1}, v_{t-1}$, where t is the current time). Given that the state of the world remains the same, there is no need to increase the count of the steps.

The fourth suggestion is: When you play an incorrect move, you to be informed that the move is incorrect but without prolongation of the history. The new history will look like this: $u_1, a_1, v_1, \ldots, u_t, a_t, v_t$. Here u_i is the set of incorrect moves at the i-th step which we have tried and we know for sure that they are incorrect. Thus the history is not prolonged, yet the information that certain moves are incorrect is

recorded in the history. Here we assume that the order in which we've tried the incorrect moves does not matter.

The fifth option, which we will discuss in this article, will be even more liberal. In the previous suggestion we have the opportunity to try whether a move is incorrect, but if it proves correct, we have already made this move. Now we will assume that we can try different moves as many as we want and we can get information whether it is correct or incorrect for each one of them. After that, we make a move, for which we already know that it is correct. We know because we've tried it. Here, the history is the same as with suggestion No. 5, except that u_i is a tuple of two sets, the first of which is from the tried incorrect moves, the second – from the tried correct moves.

There is a sixth option and it is for every step to get full information about which moves are correct and which are not. In other words, we get u_i with all the moves in it like they have been tested. However, we do not like this option because u_i may be too large, i.e. it may contain too much information. Moreover, we assume that after some training we will have a fairly accurate idea about which move is correct and which is incorrect and will not have to make many tests in order to understand this.

Conclusion

In the field of AI our first task is to say what we are looking for, and the second task is to find the thing we are looking for. This article is entirely devoted to the second task.

Articles [Dobrev, 2000, 2005, 2013] and [Legg, 2005] which are devoted to the first task, tell us that the thing we are looking for is an intelligent program that proves its intelligence by demonstrating it can understand an unknown world and cope in it to a sufficient degree.

The key element in this article is the understanding of the world itself, and more precisely – the understanding of the state of the world. We argue that if we understand the state of the world, we understand the world itself. Formally speaking, to understand the state of the world is like reducing the problem for Reinforcement learning with partial observability to the problem for Reinforcement learning with full observability. Not accidentally, almost all articles on Reinforcement learning deal with the case of full observability. This is because it is the easiest case. The only problem in this case is to overcome the combinatorial explosion. Of course, combinatorial explosion itself is a big enough problem because we get algorithms that would work if you have an endlessly fast computer and indefinitely long time for training (long in terms of the number of steps). Such algorithms are completely useless in practice and their value is only theoretical.

In this article we presented the state of the world as infinite-dimensional vector of the values of all testable states. This seems enough to understand the state of the world, but we need a finite description and we would like this description to be as simple as possible. For this purpose, we've made three additional steps. A model of the world was introduced as a finite state machine. Each such machine

describes an infinite number of testable states and this is a very simple description which is easy to operate.

The second step was the assumption that testable states are inert and change only if specific events occur. That is, if between two checks, none of these events has occurred, we can assume that the value of the testable state has not changed.

The third step was the introduction of agents which under certain conditions may alter the values of testable states. This step is particularly important because the agent hides in itself much of the complexity of the world and thus the world becomes much simpler and more understandable. We will not describe the agent as a system of formal rules. Instead, we will approximate it with assumptions such as that it is our ally and tries to help us or that it is an opponent and tries to counteract.

Without these three steps the understanding of a more complex world would be completely impossible. Formally speaking, testable states only would be sufficient to understand the state of the world. Even testable states which prerequisites are simple conjunction dependent only on the last few steps of the past are sufficient. However, without the additional generalizations made, we would face an enormous combinatorial explosion.

Bibliography

[Dobrev, 2000] Dobrev D. AI – What is this. PC Magazine – Bulgaria, November'2000, pp.12-13. http://www.dobrev.com/AI/definition.html

[Dobrev, 2005] Dobrev D. Formal Definition of Artificial Intelligence, International Journal "Information Theories & Applications", vol.12, Number 3, 2005, pp.277-285. http://www.dobrev.com/AI/

[Dobrev, 2013] Dobrev D. Comparison between the two definitions of AI. arXiv:1302.0216, January, 2013. http://www.dobrev.com/AI/

[Legg, 2005] S. Legg and M. Hutter. A Universal Measure of Intelligence for Artificial Agents, Proc. 21st International Joint Conf. on Artificial Intelligence (IJCAI-2005), pages 1509–1510, Edinburgh, 2005.

Authors' Information

Dimiter Dobrev *– Institute of Mathematics and Informatics, Bulgarian Academy of Sciences. Acad. Georgi Bonchev Str., Block 8, 1113 Sofia, Bulgaria; e-mail: d@dobrev.com*

Major Fields of Scientific Research: Artificial Intelligence, Logic Programming