HEXAPAWN, A MACHINE IN BOXES MATCHES THAT LEARN TO PLAY

Reinforcement learning

L

A machine will, by performing elementary tasks in a totally mechanical way, manage to find the strategy of a two-player game and then win every time.

KEYWORDS: unplugged activity/learning
EDUCATIONAL OBJECTIVES: to understand the concept of reinforcement learning, to experiment with an unplugged AI, to discover related concepts: states, possibility tree, optimal algorithm, random.
SKILLS USED: understanding rules, developing a strategy, anticipating an outcome

DEVICE

Age group: from 11 years old (or younger)

Group size : 2 people (one is the human, the other the machine). The games can be played fast enough to animate a group of 10 people.

Materials needed:

- 10 green beads, 14 red beads, 13 blue beads, 18 yellow beads
- 3 white pieces, 3 black pieces
- 24 matchboxes
- glue (a stick is fine).
- a large storage box.

Preparation :

- Cut out the rectangles with the shapes and stick them on the top of the matchboxes
- Put the coloured beads corresponding to the arrows in each box
- Pplace the boxes in 3 rows (game turn 2, 4 and 6)
- Cut out and glue, or draw the 3X3 board and place 3 pieces for each player (two colours)

Duration of the activity: 15 - 30 minutes

HOW IT WORKS

1) Pieces may only move one square in a straight line.
2) Pieces can only eat the opponent diagonally.
3) There are three ways to win at Hexapawn:
   - take one of your checkers to the opposite line in the opponent's camp.
   - take all the opponent's pieces.
   - blocking the opponent (who can no longer take or advance in turn).

How to play a game

- The facilitator plays the role of the AI. He has the black pieces.
- The human player has the white pieces, he starts the game and will play as he wishes. The facilitator plays for the AI using the matchboxes which represent states of the game. Equivalent positions from the point of view of the AI's actions are not represented (for example, only the box with the left pawn advanced is represented, this is the equivalent of the situation with the right pawn advanced).

Round 1: The human player advances a piece.

Round 2: The AI plays with one of the boxes marked 2. The facilitator takes the corresponding matchbox and draws a bead at random. He plays the move of the indicated colour and keeps the pearl (e.g. puts it on the box).

Round 3: There are two possible situations: the human player wins or the game continues.

- If the human player wins, the AI considers that his move was bad, the pearl is removed from the game, this is how the AI learns!
- If the game continues, the AI considers that the move was neither good nor bad, the pearl is put back in the box.

Round 4: The AI plays with one of the boxes marked 4. The facilitator draws a random bead and plays for the AI. There are only 2 possible outcomes:
- the AI wins, in this case, the AI considers that its move was good, we put the pearl back in the box.
- the game can continue. The AI considers that his move was neither good nor bad, we keep the pearl (he puts it on the box for example).

Round 5: This round has the same game mechanics as round 3. If the AI loses, the bead from round 4 is removed from the game unless it is the last bead. The other beads are put back in their boxes. Aso....
As the game progresses, the beads corresponding to bad moves are removed, the AI learns and keeps only the moves that make it win.

TO GO FURTHER

In principle, all solved games lend themselves well to demonstrations of reinforcement learning. There are variants for Nîm (which will allow you to recycle your matches) or Morpion (but you will need 304 boxes of matches). The implementation in a program then becomes more interesting and you can find examples in Scratch or Python.

In the latter case, a few conditional instructions will allow the algorithm to explore all the possible options on its own, playing against itself. This allows students to code their AI themselves and to see the optimal strategy emerge after a few hundred games, which is a few seconds for the computer.
Finally, this game can also be used to develop strategies against the machine, such as taking risks that would seem illogical against a human, in order to explore scenarios in which the AI has not yet learned. One can also wonder about the teaching or learning method to be adopted, since instead of "punishing" by removing the undesirable pearl, one could as well "reward" by adding, for example, a pearl of the same colour to each winning move by the AI. This process has the advantage of improving the AI's performance from a statistical point of view, but it still leaves an opportunity for the human player to win.