Fiche pédagogique

Activité : Machine IA

Objectifs pédagogiques : Comprendre le principe d'un algorithme d'apprentissage par renforcement.

Notions abordées : Apprentissage par renforcement. Tirage/stratégie aléatoire. Intelligence artificielle.

Matériel nécessaire : La machine IA et ses accessoires. (La machine I. A. peut être reproduite avec des boites et des jetons de différentes couleurs.)

Niveau : A partir du cycle 4.

Déroulement : Lors de cette activité, on déroule pas à pas un algorithme d'apprentissage par renforcement pour obtenir une stratégie au jeu des bâtonnets. Voir Activité Jeu des bâtonnets..

Suivant le temps et l'envie, on peut soit d'abord faire l'Activité Jeu des bâtonnets pour que les participants trouvent la stratégie gagnante et ensuite passer à celle-ci pour expliquer comment une machine peut trouver la stratégie en faisant de l'apprentissage par renforcement, soit faire cette activité sans avoir fait l'Activité Jeu des bâtonnets au préalable et expliciter la stratégie gagnante quand on découvre le résultat de l'apprentissage.

Dans un premier temps, on considère la variante $k = 2$ (chaque joueur peut ôter 1 ou 2 bâtonnets) et $n = 8$ (8 bâtonnets au départ). La machine a 8 boîtes numérotées de 1 à 8 correspondant aux positions de 1 à 8 bâtonnets. On a également des boules de deux couleurs, disons rouge et bleu. La couleur rouge va correspondre au coup "ôter un bâtonnet" et la couleur bleue au coup "ôter deux bâtonnets". Au départ, on met autant de boules rouges que de boules bleues (disons 4 de chaque couleur) dans chaque case, mis à part dans la première boîte ou on ne met que des boules rouges car on ne peut pas ôter deux bâtonnets quand il n'en reste plus qu'un.

On fait alors jouer la machine contre elle-même de la manière (semi-alétoire) suivante. Les Joueurs 1 et 2 jouent alternativement comme suit en piochant dans les boîtes, le Joueur 1 commençant à piocher dans la boîte 8.

- Un joueur pioche dans la boîte i. S'il sort une boule rouge correspondant à "ôter un bâtonnet" , l'autre joueur va piocher dans la boîte $i - 1$. S'il sort une boule bleue correspondant à "ôter deux bâtonnets", l'autre joueur va piocher dans la boîte $i - 2$. Ainsi au début, quand il y a 8 bâtonnets, le premier joueur pioche dans la boîte 8. S'il prend une boule rouge (resp. bleue), il retire un bâtonnet. Il n'en restera que 7 (resp. 6) et le Joueur 2 ira piocher dans la boîte 7 (resp. boîte 6).
- Le premier joueur qui ne peut pas piocher a perdu. Autrement dit celui qui enlève le dernier bâtonnet a gagné.

On procède ensuite à la phase de renforcement. Le joueur qui a perdu ne remet pas ses boules dans les boîtes. A l'inverse, le joueur qui a gagné remet dans chaque boîte dans laquelle il a pioché une boule, deux boules de la couleur de celle tirée. (S'il a tiré une boule bleue (resp. rouge) de la boîte i, alors il remet deux boules bleues (resp. rouges) dans la boîte i).

Remarque : Il se peut qu'une boîte se vide. Dans ce cas, impossible de piocher dedans. On ôte alors le nombre de boules (1 ou 2) que l'on veut. On peut aussi remettre une boule rouge et une boule bleue dans la machine et faire un tirage.

En pratique, l'animateur peut expliquer la procédure en endossant le rôle des deux joueurs pendant une partie. Ensuite, il choisit successivement plusieurs paires de participants à qui il fait faire une partie chacun. On insiste alors sur le fait que les deux joueurs ne réfléchissent pas. Ils ne font que tirer des boules et retirer le nombre de bâtonnets indiqué par la couleur de la boule. Ainsi ils peuvent jouer de mauvais coups par exemple si un joueur tire une boule rouge alors qu'il reste deux bâtonnets, il va retirer un seul bâtonnet laissant le joueur adverse retirer le dernier et gagner

On effectue plusieurs fois la procédure (jeu + renforcement) et on observe ce qui se passe. L'idée est de faire observer les participants et de les amener à voir que les boîtes n'ont plus 4 boules rouges et 4 boules bleues et qu'il y a différents types de boîtes. En effet, avec une probabilité (qui devient de plus en plus forte à chaque qu'on refait la procédure), on voit la stratégie gagnante apparaitre : les boîtes correspondant aux positions gagnantes ont une très large majorité de boules de la couleur correspondant au coup gagnant dans cette position. Les positions perdantes se vident. Ainsi pour le jeu où on enlèvent un ou deux bâtonnets, la boîte 1 ne va avoir que des boules rouges, la boîte 2 que des boules bleues, la boîte 3 se vide, la boite 4 n'aura que des boules rouges, la boîte 5 que des boules bleues, la boîte 6 se vide, etc .... Cela correspond bien à la stratégie gagnante : les positions perdantes sont les multiples de 3 ; quand il y a un multiple de 3 plus 1 bâtonnets il faut en retirer 1 ; quand il y a un multiple de 3 plus 2 bâtonnets il faut en retirer 2.

Il est possible d'expliquer que la machine va forcément converger vers la configuration décrite ci-dessous, en regardant ce qui se passe pour les premières les premières boîtes. La boîte 1 n'a que des boules rouges et cela va continuer. Regardons la boîte 2. Si on y pioche une boule bleue, alors on retire deux bâtonnets et donc on gagne. On va donc remettre la boule bleue ainsi qu'une boule bleue supplémentaire. Si on y pioche une boule rouge, alors on retire 1 bâtonnets et le joueur adverse va retirer le dernier bâtonnet. On perdra donc et la boule rouge sera ainsi retirée. Donc, dans la boîte 2, les boules rouges disparaissent et les bleues se multiplient. Au bout d'un moment, il n'y aura donc que des bleues. On peut alors voir que la boîte 3 se videra, car qu'on tire une boule rouge ou une boule bleue, l'adversaire gagnera et donc on devra ôter la boule de la boîte. Et ainsi de suite.

On peut faire le processus plusieurs fois pour observer que les résultats concordent et pour différentes variantes du jeu. Cela peut se faire rapidemement en utilisant l'application Java créée par E. Duchêne et A. Parreau (MMI, Univ. Lyon 1, LIRIS) téléchargeable à la page :
`https://projet.liris.cnrs.fr/lirismed/index.php?id=la-machine-qui-apprend-a-jouer-toute-seule`

Il est important également de faire remarquer les choses suivantes qui permettent d'ouvrir

ensuite sur une discussion sur les algorithmes d'apprentissage et la différence entre apprentissage humain et apprentissage machine.

1. La stratégie rendue par un algorithme d'apprentissage par renforcement est une stratégie probabiliste. En général, à la fin on a une probabilité de jouer les différents coups (même si on peut la transformer en stratégie déterministe en choisissant le coup le plus probable). Et contrairement à ce que peut faire un humain, la machine ne fournit pas de preuve que la stratégie est gagnante.
2. Il faut de très nombreuses parties pour que l'algorithme par renforcement converge vers une stratégie proche de la stratégie gagnante, et ce d'autant plus que le jeu est complexe (multiples positions et de nombreux coups possibles). Pour le jeu des bâtonnets, un humain peut trouver la stratégie en ne faisant qu'au plus quelques dizaines de parties, là où une machine aura besoin de quelques centaines de parties. Un humain apprend donc en beaucoup moins de parties qu'une machine, mais une machine fait des parties beaucoup plus rapidement qu'un humain et apprend donc plus rapidement. Cependant, l'humain peut parfois cependant généraliser sa stratégie. C'est le cas pour le jeu des bâtonnets, avec la stratégie décrite ci-dessus (en fonction des multiples de 3). Ainsi l'humain a la stratégie quel que soit le nombre de bâtonnets, là où une machine devrait recommencer ou continuer à jouer pour un plus grand nombre.
3. Le fait que la machine ne puissent pas généraliser ou identifier des motifs fait que les algorithmes obtenus par apprentissage sont le plus souvent plus gros que ceux créés par les humains. Par exemple, pour le jeux des bâtonnets où chaque joueur ôte un ou deux bâtonnets, un humain fera un algorithme avec les trois opérations décrites ci-dessus, là où une machine aura une action par nombre de bâtonnets restants.
4. L'avantage d'un algorithme d'apprentissage est qu'on a pas besoin de réfléchir pour trouver une stratégie gagnante ou gagnant souvent. Il donne de très bonnes stratégies (mais qu'on en sait pas montrer gagnantes le plus souvent) y compris pour des jeux compliqués tels que les échecs pour lesquels l'existence d'une stratégie gagnante est inconnue.

En parlant d'échecs, on peut mentionner AlphaZero (variante d'AlphaGo) qui a appris à jouer aux échecs en faisant 3 milliards de parties contre elle-même (noter qu'en faisant 10 parties par jour, soit 3650 parties par an, et pendant 100 ans, un joueur ferait moins de 400 000 de parties, soit dix mille fois moins que ce qu'à fait l'ordinateur), est devenu le meilleur programme de jeux d'échecs. Le gros avantage d'un ordinateur c'est qu'il peut faire très rapidement, bien plus de parties qu'un humain. En revanche, contrairement à un joueur d'échecs confirmé, l'ordinateur ne peut pas expliquer pourquoi une position est bonne ou mauvaise. Il peut au mieux indiquer une probabilité de victoire en étant dans cette position.