

TES	APPRENTISSAGE PAR RENFORCEMENT	Activité
-----	--------------------------------	----------

Document 1 : Intelligence artificielle et jeux

Dès les débuts de l'intelligence artificielle, les chercheur·ses se sont intéressé·es aux jeux. Passons en revue quelques grandes dates.

Dès les années 1950, Arthur Samuel (figure 1), travaillant chez IBM, élabore un programme de jeu de dames (dans sa version anglaise, sans mouvement long de la dame) et l'améliorera jusqu'au début des années 1960. C'est en 1959 qu'il introduit un peu d'apprentissage par renforcement dans son programme et c'est à ce moment-là qu'apparaît le terme d'apprentissage automatique (machine learning). Le jeu de dames anglaises a beaucoup plus tard été faiblement résolu par Schaeffer en 2007 : il y a toujours une stratégie de match nul (pour les deux joueur·ses) et Schaeffer construit un programme qui assure le nul contre un·e joueur·se jouant parfaitement mais qui n'est pas forcément capable de gagner contre un·e joueur·se faisant des erreurs (le programme ne transformera jamais une situation de nul en défaite mais peut transformer une position pouvant mener à la victoire en nul - c'est le sens de faiblement résolu).

Entre-temps, on peut noter en 1997 la victoire de Deep Blue, logiciel d'IBM, contre Garry Kasparov, alors champion du monde des échecs (figure 2). Mais, aux échecs, on ne sait toujours pas si les blancs ont une stratégie de victoire (c'est-à-dire s'ils sont absolument certains de gagner en jouant optimalement), on sait juste faire des programmes bien meilleurs que les humain·es.

En 2016, AlphaGo défraie la chronique en battant l'un des meilleurs joueurs du monde au Go (figure 3). Si la victoire de la machine ne surprend pas les spécialistes qui savaient que cela finirait par arriver, c'est la rapidité à laquelle cela est arrivé qui a étonné.

Et depuis, l'intelligence artificielle brille au poker, à des jeux de société (pouvant même nécessiter tromperie ou empathie comme Diplomacy), des jeux vidéo, ...

Source : <https://culturesciencesphysique.ens-lyon.fr/ressource/IA-Nim-1.xml>

On va voir dans cette activité comment la machine peut apprendre à gagner à une variante du jeu de Nim.

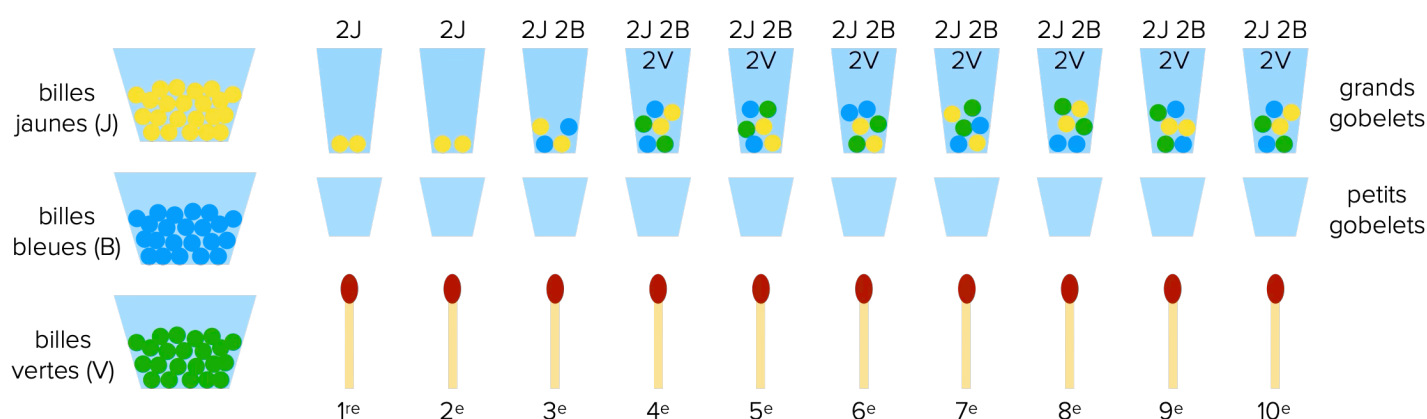
Document 2 : règles du jeu

10 allumettes sont disposées devant deux joueurs.

Tour à tour, chaque joueur peut prendre 1, 3 ou 4 allumette(s) (pas 2 et pas plus de 4).

Le joueur qui prend la ou les dernières allumettes a gagné.

On forme des équipes de 4 ou 5. Pour construire la machine, on reproduit la configuration du schéma ci-dessous.



Un des membres de l'équipe opère la machine quand un autre joue contre elle.

La machine commence.

Pour la faire jouer, l'opérateur tire au hasard une bille dans le grand gobelet situé au niveau de l'allumette la plus à droite (la 10^e) et la place dans le petit gobelet devant le grand.

- Si la bille est jaune, l'opérateur retire l'allumette la plus à droite (la 10^e).
- Si la bille est bleue, l'opérateur retire les trois allumettes les plus à droite (10^e, 9^e et 8^e).
- Si la bille est verte, l'opérateur retire les quatre allumettes les plus à droite (10^e, 9^e, 8^e et 7^e).

C'est ensuite au tour de l'humain de jouer. Il retire une, trois ou quatre des allumettes restantes les plus à droite.

C'est à nouveau au tour de la machine. L'opérateur tire maintenant au hasard une bille dans le gobelet correspondant à l'allumette la plus à droite et la place dans le petit gobelet devant. Puis il retire les allumettes en fonction de la couleur de la bille comme au premier tour.

La partie continue jusqu'à ce qu'un joueur gagne en prenant la ou les dernières allumettes.

➔ **Si la machine a perdu :**

On punit sa stratégie en retirant du jeu les billes qu'elle a tirées et qui se trouvent dans les petits gobelets (on les remet dans les trois réserves de billes situées à gauche).

➔ **Si la machine a gagné :**

On récompense sa stratégie en plaçant dans chacun des petits gobelets contenant une bille une deuxième bille de la même couleur. Puis on verse les deux billes dans le grand gobelet correspondant.

On recommence ensuite une nouvelle partie (on peut bien sûr changer d'opérateur et de joueur humain).

Il faut faire le plus de parties possibles (au moins une quinzaine).

Si un grand gobelet se retrouve vide, on le remplit comme au départ avec les billes de la réserve.

1. Au départ, que peut-on dire de la probabilité du choix du coup par la machine (entre retirer une, trois ou quatre allumettes) lorsqu'il reste entre 10 et 4 allumettes ?
2. Expliquer les remplissages initiaux des gobelets correspondant à trois allumettes restantes, deux restantes et une restante.
3. Comment les récompenses ou punitions modifient-elles le jeu de la machine ?
4. Aux vues des billes présentes dans les gobelets après un grand nombre de parties, émettez une hypothèse quant à la stratégie gagnante pour le joueur 1.

On va maintenant tâcher de déterminer à la main cette stratégie gagnante.

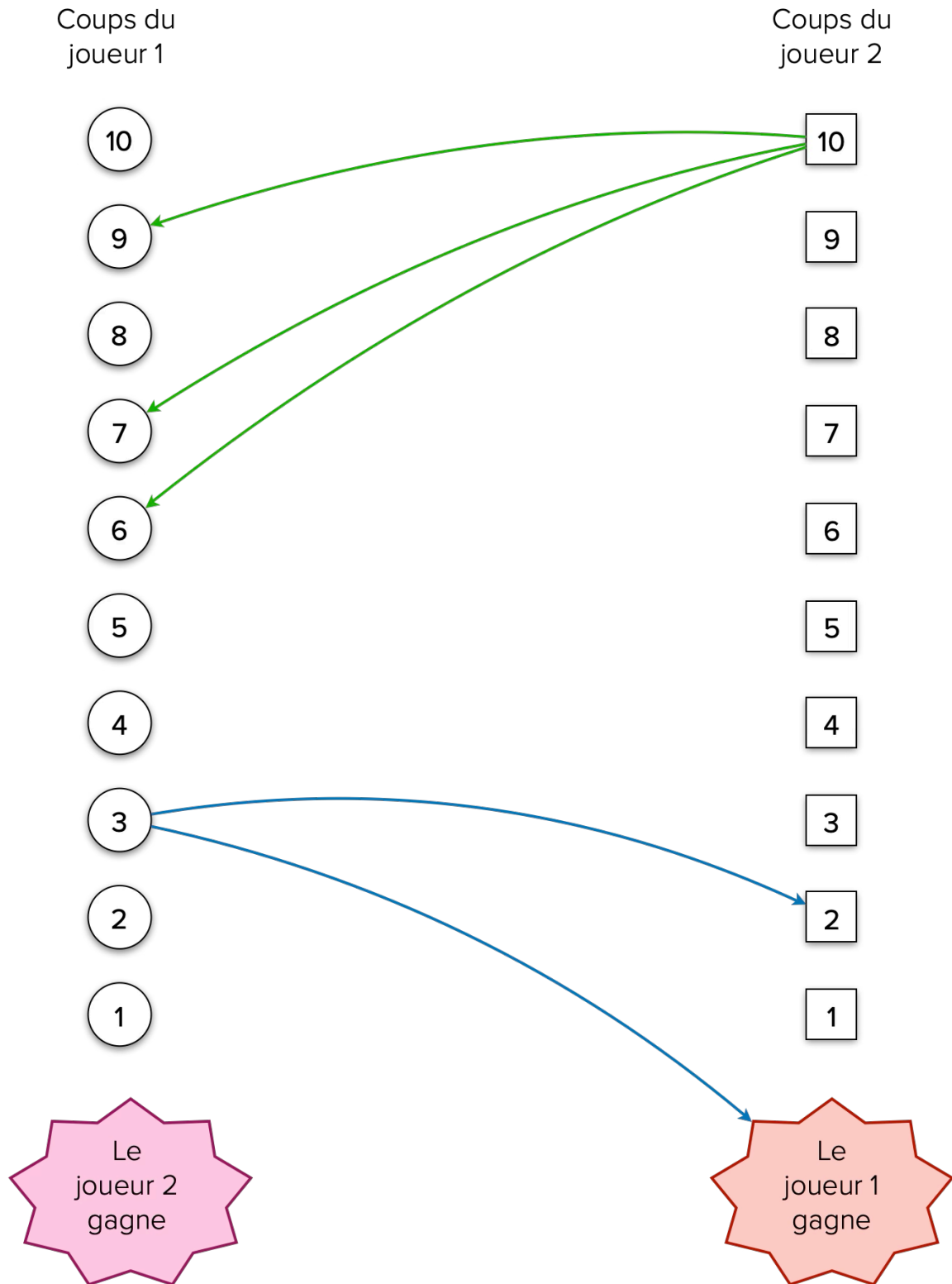
Pour cela, on va dresser le graphe du jeu qui consiste à relier d'une flèche chaque position possible d'un joueur aux positions de l'autre joueur qu'il peut atteindre.

Un exemple de position possible pour le joueur 2 est : 10 allumettes restantes. Il peut alors atteindre 3 positions du joueur 1 : 5 allumettes restantes (en retirant une allumette), 3 allumettes restantes (en retirant 3 allumettes) et 2 allumettes restantes (en retirant 4 allumettes).

Un exemple possible de position pour le joueur 1 est : 3 allumettes restantes. Il peut alors atteindre 1 position du joueur 1 : 2 allumettes restantes (en prenant une allumette). Et il peut aussi gagner la partie (en prenant les 3 allumettes).

On va représenter les positions du joueur 1 par des ronds et celles du joueur 2 par des carrés. Le chiffre dans le rond ou le carré correspond au nombre d'allumettes restantes.

Les deux exemples qu'on vient de détailler ont déjà été complétés. À vous de compléter le reste des coups possibles pour chacun des joueurs.



Pour déterminer une stratégie gagnante pour le joueur 1, on suit l'algorithme suivant :

- On part de « le joueur 1 gagne ». On colorie en rouge chacune des positions du joueur 1 qui permettent d'atteindre cette victoire (ce sont des **positions gagnantes pour le joueur 1**) et on surligne les flèches qui rejoignent la victoire.
- S'il existe des positions du joueur 2 dont **toutes les flèches** rejoignent des positions gagnantes du joueur 1, alors ces positions sont des **positions perdantes pour le joueur 2**. On colorie ces positions en bleu et on surligne les flèches qui les piègent sur des positions gagnantes pour 1.
- Si **au moins une flèche** d'une position de 1 rejoint une position perdante de 2, alors cette position est une position gagnante pour le joueur 1 (puisque'il suffit au joueur 1 de jouer ce coup pour piéger le joueur 2). On colorie alors en rouge ces nouvelles positions gagnantes du joueur 1 et on surligne les flèches les joignant à des positions perdantes pour 2.
- On continue jusqu'à ne plus pouvoir colorier de position.

Si on a fini par colorier la position 10, on a alors une stratégie gagnante pour le joueur 1 s'il commence ! Il suffit de suivre les flèches rouges.

5. Est-ce que la stratégie gagnante pour le joueur 1 déterminée grâce au graphe correspond à peu près aux billes se trouvant dans les gobelets ? Justifier.