

ATELIER

Jeu de Nim



Atelier – Jeu de Nim

Cet atelier illustre les principes de l'apprentissage par renforcement, une technique qui permet aux ordinateurs d'« apprendre » à réaliser des tâches complexes, ce qui inclut notamment la capacité à gagner à des jeux de stratégie. L'atelier permet d'expérimenter l'entraînement d'une machine qui apprend à jouer à un jeu simple, le jeu de Nim, et de comprendre les forces et les limites des algorithmes d'apprentissage.

Matériel :

- 9 bâtonnets (ou allumettes ou crayons ou cure-dents)
- 16 gobelets en carton (éventuellement numérotés de 1 à 8)
- 60 billes, jetons ou autre petit objet en 3 couleurs (au moins 20 jetons de chaque couleur)
- Une table
- Papier et crayons pour noter le nombre de parties, ou sinon pastilles, post-it ou tout autre moyen de compter

Durée : entre 30 et 45 minutes

Age du public : lycéens (15–18 ans) ou collégiens (13–14 ans)

Introductions

Informations

En bleu figurent les moments où les élèves expérimentent, en vert les moments où nous intervenons, en noir le développement et les explications de l'atelier.

Règles du jeu de Nim (à connaître avant de démarrer)

Ce jeu se joue à 2 joueurs. 9 bâtonnets sont posés sur une table : à chaque tour, chaque joueur retire 1 ou 2 bâtonnets de la table, jusqu'à ce qu'il n'en reste plus aucun. Le joueur qui se trouve à prendre le dernier bâtonnet a perdu. On peut choisir aussi de jouer suivant une autre règle pour le retrait des bâtonnets : le choix est alors entre en retirer 1, 3 ou 4.

Le jeu de Nim est très ancien et il en existe plusieurs versions, on parle plutôt de jeux de Nim. Le nombre total de bâtonnets peut varier ainsi que le nombre que l'on peut retirer à chaque coup ; on peut aussi décider que le joueur qui retire le dernier bâtonnet est le gagnant et non pas le perdant. Dans tous les cas, il s'agit d'un jeu de **stratégie pure**, où la victoire ne dépend pas du hasard et où aucune partie ne peut se terminer par une égalité : il existe donc une stratégie gagnante pour l'un des deux joueurs.

Étape 1 : préparation du matériel.

Disposer les 9 bâtonnets les uns à côté des autres sur une table. Face à chaque bâtonnet, sauf le dernier, poser un gobelet : dans chaque gobelet mettre 4 jetons de 2 couleurs différentes, par exemple rouge et bleu.



Figure 1. Configuration initiale du jeu de Nim.

Étape 2 : première version du jeu.

Deux joueurs s'affrontent au jeu de Nim : un représente un être humain, l'autre un ordinateur. Le joueur « ordinateur » commence la partie. Elle ou il prend le premier gobelet, pioche un jeton au hasard et le pose devant le gobelet. En fonction de la couleur du jeton, elle ou il enlève un certain nombre de bâtonnets. Disons par exemple qu'en cas de jeton rouge on



retire 1 bâtonnet et en cas de jeton bleu on en retire 2. Le joueur « humain » joue à son tour en retirant le nombre de bâtonnets qu'elle ou il souhaite. Le joueur « ordinateur » prend alors le gobelet qui se trouve face à la première des allumettes restantes, pioche un jeton et ainsi de suite. Le jeu continue jusqu'à ce que l'un des deux gagne. À la fin de la partie, si l'ordinateur a perdu, on met les jetons extraits dans une défausse, pour qu'ils ne se retrouvent plus dans les gobelets. Si au contraire il a gagné, on remet chaque jeton dans son gobelet et on rajoute en plus un jeton de la même couleur pour chaque jeton remis dans son gobelet. On joue plusieurs parties suivant ces règles, en alternant entre une partie où l'humain commence et une partie où la machine commence. Si le groupe est nombreux, les élèves peuvent alterner pour jouer le rôle de l'humain ou de la machine. Si un gobelet reste vide, on le "réinitialise" en remettant les 4 jetons qu'il y avait au départ.

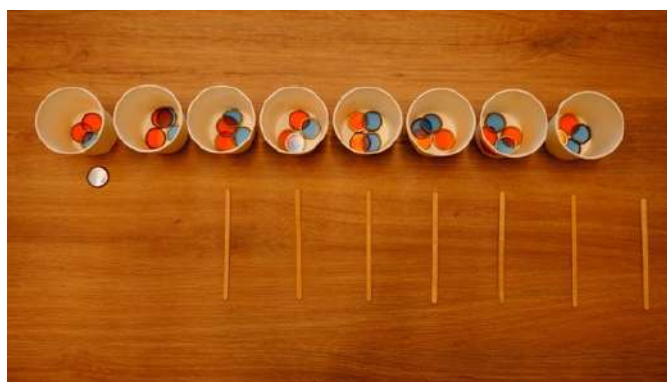


Figure 2. Au premier tour, le joueur « ordinateur » a pioché un jeton bleu. Elle ou il retire alors 2 bâtonnets et pose le jeton devant le gobelet correspondant.

Au bout d'une dizaine de parties, on regarde le contenu des gobelets : qu'est-ce qu'on remarque ? Qu'est-ce qui a changé par rapport au contenu initial ? Discussion.

Pour réaliser une machine imbattable, il y a maintenant deux choix : continuer à jouer des parties jusqu'à terminer l'entraînement de l'ordinateur, ou trouver la bonne stratégie par le raisonnement. On laisse éventuellement le groupe jouer quelques autres parties et on guide les participants et participantes vers la détermination de la solution. La meilleure façon de procéder est de trouver le coup gagnant pour la dernière position, puis pour l'avant-dernière et ainsi de suite jusqu'à la première.

Lorsqu'on trouve le coup gagnant pour une certaine position, on remplace le gobelet par un simple jeton de la bonne couleur. Si la position est perdante quoi que l'on fasse, on laisse un gobelet vide pour que l'ordinateur arrête de jouer quand il atteint cette position.

Étape 3 : deuxième version du jeu.

On recommence le jeu avec des nouvelles règles : au lieu de choisir entre 1 ou 2 bâtonnets à retirer à chaque tour, le choix est entre 1, 3 ou 4. Il faut alors mettre des jetons de 3 couleurs différentes dans les gobelets : par exemple rouge pour 1 bâtonnet, bleu pour 3 et jaune pour 4.



Figure 3. Configuration initiale du jeu dans sa deuxième version.

Les élèves jouent quelques parties suivant ces nouvelles règles.

De nouveau, on regarde le contenu des gobelets au bout de quelques parties et on discute pour trouver la meilleure stratégie. On vérifie s'il y a bien une stratégie en train d'émerger à travers l'entraînement et si elle est la gagnante.

Discussion : quelle est la meilleure façon de jouer de la part de l'humain pour que l'ordinateur apprenne rapidement la bonne stratégie ? Faut-il faire en sorte que la machine gagne souvent, ou qu'elle perde souvent ? Jouer au hasard ou suivre un schéma dans ses parties ?

Étape 4 : troisième version du jeu (optionnelle).

On fait maintenant jouer deux machines l'une contre l'autre. Pour ce faire, on dispose deux files de gobelets remplis de jetons, une de chaque côté de la file de bâtonnets. On joue avec la règle de 1 ou 2 bâtonnets, donc avec 2 couleurs de jetons.



Figure 4. Configuration initiale du jeu dans la version «machine contre machine».

Les élèves jouent des parties dans cette nouvelle configuration. Les machines s'alternent dans le rôle de celle qui commence la première, suivant toujours les mêmes règles.

Au bout de quelques parties, on vérifie le contenu des gobelets et on échange sur la stratégie qui est en train d'émerger. Ce n'est pas grave si on n'a pas le temps de jouer beaucoup de parties, l'important ici est d'illustrer le principe.

Principe de l'apprentissage par renforcement.

Dans toutes les versions du jeu qui auront été jouées, on applique toujours le même **algorithme** pour faire apprendre la machine. Lorsqu'on joue un certain coup et que l'on perd la partie, on retire le jeton correspondant. Dans le gobelet il y aura donc un jeton de moins pour une certaine couleur, ce qui réduit la probabilité de piocher cette couleur et donc de jouer à nouveau le même coup. Vice-versa, si en jouant un certain coup on a gagné, on rajoute un deuxième jeton de la même couleur pour **renforcer** la probabilité de gagner à nouveau. Par conséquent, au bout de quelques parties, certains gobelets se retrouveront remplis de jetons de la même couleur. Par exemple : si c'est à l'ordinateur de jouer lorsqu'il reste deux bâtonnets sur la table, prendre les deux conduit à une défaite et prendre un seul conduit à la victoire. Le huitième gobelet finira donc pour ne contenir que des jetons rouges. Certains gobelets se retrouveront vides : ceci veut dire qu'il s'agit d'une position perdante quoi que l'on fasse.

Jeux déterministes.

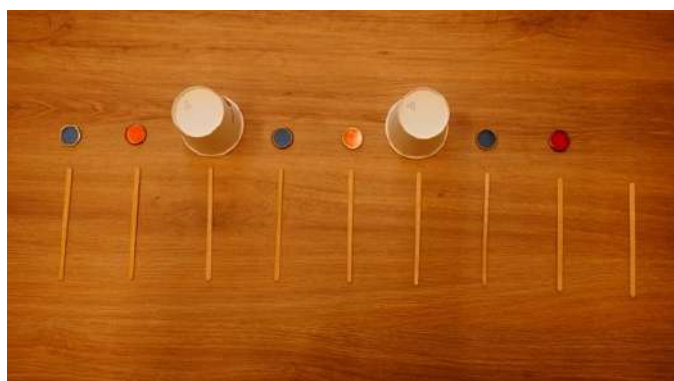


Figure 5. Solution du jeu dans la première configuration.

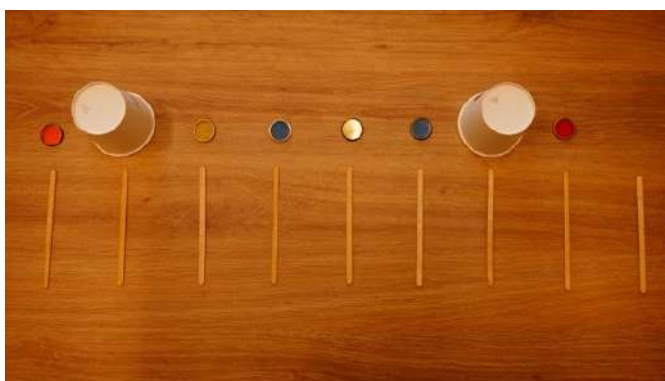


Figure 6. Solution du jeu dans la deuxième configuration.

La stratégie gagnante pour la première version du jeu est **2-1-X-2-1-X-2-1-X** (2 = retirer 2 bâtonnets, 1 = retirer 1 bâtonnet, X = perdu). Pour la deuxième version, la stratégie est **1-X-4-3-4-3-X-1-X**. Le jeu de Nim est complètement déterministe, dans le sens que la victoire ne dépend pas du tout du hasard : une fois que l'on connaît la stratégie gagnante pour une position de départ favorable, on peut toujours gagner lorsqu'on est le joueur ou la joueuse qui commence la partie. Ceci s'applique pour toutes les versions du jeu, peu importe le nombre de bâtonnets et les règles pour les retirer. S'il y avait une composante de hasard, l'algorithme d'apprentissage par renforcement fonctionnerait moins bien, car les mêmes coups conduiraient parfois à la victoire et parfois à la défaite.

Les versions du jeu de Nim présentées dans cet atelier sont suffisamment simples pour que l'on trouve la stratégie gagnante par le raisonnement, ce qui est beaucoup plus rapide que jouer plusieurs parties. Ceci permet aussi de vérifier si l'algorithme d'apprentissage fonctionne,

en regardant la couleur des jetons qui se retrouvent dans les gobelets au bout de quelques parties. Pour des jeux plus complexes comme la dame, le go ou même le morpion, où il y a beaucoup plus de coups possibles, trouver les stratégies gagnantes est loin d'être évident : il a fallu environ 7 ans pour que le AlphaGo de Google – DeepMind batte le champion du monde du jeu de Go Ke Jie (Deep Mind a été fondée en 2010, rachetée par Google en 2014 et obtenu la victoire contre Ke Jie en 2017). Il faut savoir par contre qu'AlphaGo utilisait plutôt une combinaison d'algorithmes, notamment l'apprentissage par renforcement et la recherche arborescente, qui permet de tester différents types de parties suivant une méthode précise.

Est-ce que l'apprentissage par renforcement a rendu la machine intelligente ?

Eh bien, non ! La machine n'a pas « appris » le raisonnement qui conduit à la victoire, elle a juste testé des manières de réaliser la tâche « gagner au jeu de Nim » et a retenu les manières qui fonctionnent : c'est nous d'ailleurs qui avons « communiqué » à la machine quand elle a réussi et quand non. Si les règles changent, comme cela a été le cas entre la première et la deuxième version du jeu, il faut recommencer l'entraînement de zéro.

Dans les versions du jeu de Nim que l'on a jouées, il n'y aurait pas besoin d'utiliser l'apprentissage par renforcement pour créer une machine imbattable : il suffirait de lui faire mémoriser la solution qu'on a trouvé par le raisonnement. Par contre, l'apprentissage par renforcement devient très utile si on veut devenir imbattable à un jeu plus complexe, là où une stratégie gagnante est beaucoup plus longue et complexe à trouver, car un ordinateur peut tester un grand nombre de parties beaucoup plus rapidement qu'un humain. Le même principe s'applique dans l'apprentissage d'autres tâches que la victoire à un jeu, comme par exemple la reconnaissance de voix ou d'images.

Quelle est la meilleure manière de jouer pour que la machine trouve rapidement la bonne stratégie ?

Si on laisse la machine gagner trop souvent, elle finit par apprendre des stratégies qui ne mènent pas à la victoire. À l'inverse, si elle perd trop souvent, elle ne sera pas en mesure d'apprendre car il n'y aura aucun renforcement. Il faut donc qu'elle alterne entre défaites et victoires.

Il y a un nombre fini de combinaisons de coups à tester, chacune correspondant à une succession de choix des deux joueurs. Dans la première version, on a $2^8=256$ possibilités (2 choix par emplacement, 8 emplacements), dans la deuxième version $3^8=6561$. Pour que l'ordinateur devienne imbattable au jeu, il faut qu'il teste tous les coups et qu'il ait un retour sur le résultat (victoire ou défaite). L'humain doit donc éviter de jouer tout le temps de la même manière (par exemple, en prenant toujours un bâtonnet la première fois), sinon la machine apprendra certaines parties mais pas d'autres. Si par contre on change de stratégie trop souvent, la machine mettra plus longtemps pour trouver les parties gagnantes. Le mieux est donc que l'humain joue suivant son inspiration, sans trop réfléchir mais sans faire exprès de

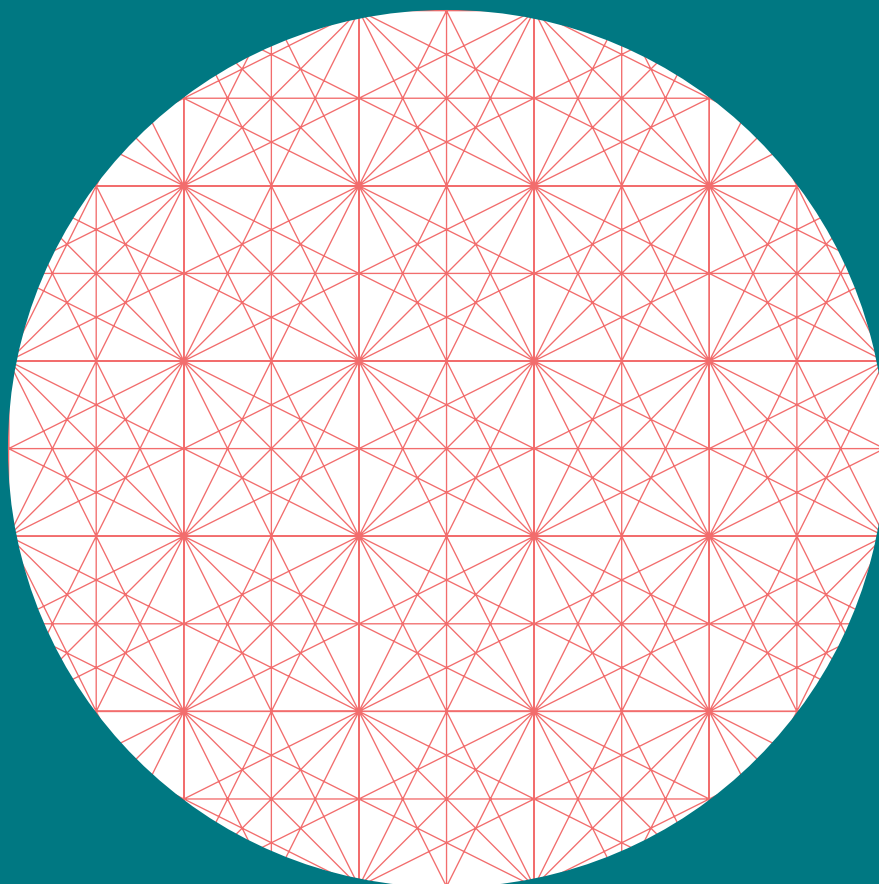


perdre non plus, en testant plusieurs façons de jouer. Mieux encore est d'éliminer le facteur humain et de mettre la machine devant une entité qui joue comme elle : une autre machine. C'est en la faisant jouer contre elle-même qu'AlphaGo a été améliorée, devenant plus forte en encore moins de temps.

Modalités d'animation :

L'atelier peut être réalisé avec des groupes allant de 2 à 35 élèves. Les groupes nombreux peuvent être divisés en sous-groupes de 6-8 élèves (ou moins), chacun avec son matériel : il faut alors multiplier le matériel listé en haut par le nombre de groupes. Dans chaque groupe, les élèves jouent des parties en jouant à tour de rôle l'humain ou la machine. En fonction du temps disponible, on peut laisser les participants et participantes réaliser plus ou moins de parties dans chaque configuration de jeu (1-2 bâtonnets, 1-3-4 bâtonnets, machine contre machine), sachant que chaque configuration nécessite de plus de temps d'entraînement que la précédente.

S'il y a suffisamment de temps à disposition, on peut tenir trace du nombre de parties nécessaire pour que l'ordinateur devienne imbattable, afin de comparer les cas de machine contre humain et de machine contre machine.



Institut Henri Poincaré | Maison Poincaré
11, rue Pierre et Marie Curie, 75005 Paris

profs-maison-poincare@ihp.fr

01 44 27 64 73

maison-poincare.fr

