

## Intelligence artificielle et jeu de Nim

### Notion d'apprentissage automatique *machine learning*

L'intelligence artificielle contemporaine et ses succès récents (depuis 2000-2010) reposent en grande partie sur **l'apprentissage machine**. Celui-ci est utile lorsqu'on ne sait pas comment décomposer une tâche compliquée en sous-tâches simples. La programmation classique consiste à donner une suite d'instructions simples permettant d'arriver au but recherché. Parfois, nous sommes dans l'incapacité de le faire, soit tout simplement parce que nous, humains, ne savons pas comment faire, soit parce que nous savons parfaitement effectuer la tâche mais sommes incapables de la décomposer en éléments simples. Dans ce cas, une piste est d'utiliser de l'apprentissage. Cela reste de la programmation. Simplement, on explique comment apprendre à arriver au but sans expliquer directement comment y arriver.

On peut distinguer trois grands types d'apprentissage aujourd'hui :

- **l'apprentissage supervisé** utilisé par exemple pour la reconnaissance d'images. L'idée est que la machine apprenne à partir d'un jeu de données correctement étiqueté suffisamment large pour que les prédictions soient bonnes par la suite.
- **l'apprentissage non supervisé** utilisé par exemple dans les systèmes de recommandation. Dans ce cas les données ne sont pas classifiées. C'est l'algorithme qui va trouver des motifs en trouvant des similarités dans les données et créer des classes.
- **l'apprentissage par renforcement** utilisé dans les prises de décision. L'algorithme apprend par l'expérience. Pour chaque tâche, l'utilisateur humain ou une machine lui donne une récompense positive ou négative.

Dans la plupart des applications, les trois sont mélangés.

Dans la suite, nous proposons une activité permettant d'éclaircir ce concept d'apprentissage en se concentrant sur l'apprentissage par renforcement.

### Apprentissage automatique et jeux

- En 1959, Arthur Samuel, travaillant chez IBM, met au point dans son programme de jeu de Dames un peu d'apprentissage par renforcement. C'est là que ce créer le terme *machine learning* ou apprentissage automatique en français.
- 1997 la victoire de Deep Blue, logiciel d'IBM, contre Garry Kasparov, alors champion du monde des échecs. Le programme permettait de calculer les possibilités jusqu'à 15 coups en avance. Connaissait de nombreux début de partie et de fin de partie et avait accès à des nombreuses parties célèbres.
- En 2016, AlphaGo de DeepMind propriété de Google bat l'un des meilleurs joueurs du monde au Go.

D'autres jeux sont explorés et des questions restent sans réponse. En effet, on ne sait toujours pas si au échec les blancs ont une stratégie gagnante. C'est-à-dire

qu'elle permet de gagner de façon certaine, les programmes jouent juste mieux que les humains.

## Le jeu de Nim

### Découverte du jeu



### Règle

Poser les 20 allumettes devant vous comme sur la photo.

- Chaque joueur prend 1 ou 2 allumettes
- Le joueur qui prend la ou les dernières allumettes a gagné.

### Question:

Quelle est la stratégie gagnante pour ce jeu?

### Apprendre à une machine à jeu à ce jeu

Pour accélérer l'apprentissage, nous allons utiliser que 8 allumettes.

Le choix de la machine sera représenté par:

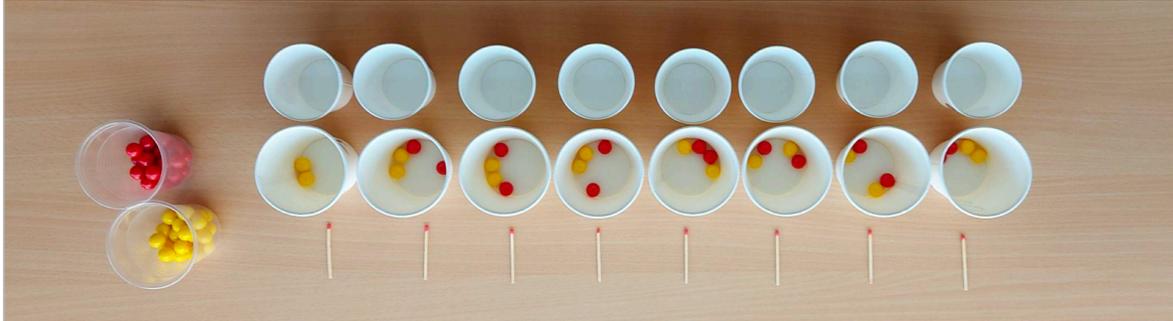
- Une bille jaune pour retirer une allumette;
- Une bille rouge pour retirer deux allumette.

**Matériel** Vous aurez besoin, pour chaque machine, de :

- huit verres en carton ou plastique numérotés de 1 à 8
- huit petits gobelets ou récipients
- des billes de couleur (jaune et rouge ici)

- huit allumettes pour jouer contre la machine
- deux verres de réserve de billes.

**Installation de la machine** Il faut placer les huit verres numérotés de 1 à 8 avec les huit allumettes en vis-à-vis ainsi qu'un petit récipient devant chaque verre. Dans les verres 2 à 7, il faut placer deux billes de chaque couleur (jaune et rouge pour nous). Dans le verre 1, il faut mettre seulement deux billes jaunes, car lorsqu'il ne reste qu'une allumette le seul choix possible de la machine est de retirer une allumette.



### Faire jouer la machine

Il faut deux personnes:

Un joueur et un opérateur qui va faire jouer la machine. Afin de vérifier qu'il n'y a pas d'erreur commise, il va y avoir aussi deux observateurs qui s'assurent du bon fonctionnement du jeu. À chaque partie le rôle de chacun peut changer.

- La machine commence à jouer. L'opérateur prend une bille au hasard dans le gobelet numéroté 8. La bille tirée est mise dans le récipient devant.
  - Si la bille est jaune, l'opérateur retire une allumette
  - Si la bille est rouge, l'opérateur retire deux allumettes
- Le joueur retire une ou deux allumettes.
- C'est à nouveau au tour de la machine de jouer. L'opérateur tire au hasard une bille dans le gobelet correspondant au nombre d'allumettes restant et enlève le nombre d'allumettes correspondant.
- Ce processus est recommencé jusqu'à la fin du jeu.

### Phase d'apprentissage de la machine

À la fin du jeu:

- Si la machine a perdu, on va simplement ne pas remettre les billes jouées (stockées dans les récipients) dans les verres dont elles proviennent. On les remet dans la réserve. C'est une punition (terme consacré en apprentissage machine). Cela correspond à un mauvais choix dans la base des possibilités de jeu.
- Si la machine a gagné, on va remettre ces billes dans les verres en ajoutant une (prise de la réserve) de la même couleur. C'est une récompense. Par exemple, si la machine a joué rouge au premier coup et qu'elle a gagné la partie, on va remettre la bille rouge plus une autre issue de la réserve

dans le verre 8. On fait de même avec tous les verres dans lesquels la machine a pioché au cours de la partie.

L'idée de l'apprentissage est la suivante : si la machine a perdu, la succession de coups qu'elle a joués a eu une issue malheureuse et on a envie qu'elle les refasse moins souvent. C'est pourquoi on ne remet pas les billes dans les verres, afin de diminuer les probabilités de rejouer ces coups. Par contre, si elle a gagné, cette succession de coups a une issue heureuse et on souhaite qu'elle les rejoue plus souvent, c'est pourquoi on augmente les probabilités qu'elle rejoue ces coups en ajoutant une bille de la couleur jouée dans le verre. C'est bien de l'apprentissage par renforcement car on renforce la probabilités des choix permettant de gagner et amoindri celle de perdre.

Dans cette machine, l'opérateur joue une fonction tirage au sort implémentable en programmation, les petits gobelets retournés servent à stocker en mémoire les coups joués pendant la partie, les grands verres contenant les billes stockent en mémoire les probabilités de jouer une ou deux allumettes au cours de l'apprentissage.

Voici le programme correspondant:

<https://nsi1.frama.io/basthon-console/?from=/algo/algo2/jeudeNim-humain.py>

Les gobelets sont représentés par des tableaux, une partie est représentée par un tableau. À la fin de chaque partie, le contenu des tableaux est mis à jour en selon les règles appliquées lors du jeu avec les gobelts.

## Analyse du jeu

Tout d'abord, analysons le jeu. Ce jeu est un **jeu à deux**, à **information complète** (on connaît à tout moment les coups possibles présents et futurs des deux joueurs), **sans hasard**, se terminant en un **nombre fini de coups** (chaque coup présentant lui-même un nombre fini de possibilités), **avec victoire** d'un des deux joueurs à la fin (il n'y pas de partie nulle). Pour un tel jeu, une des deux joueurs (que nous appellerons maintenant J1 et J2) a une stratégie gagnante, c'est-à-dire qu'il peut jouer une série de coups (dépendant de ceux de son adversaire) qui, quoique fasse son adversaire, le mènera à la victoire.

Dans le cas du jeu des allumettes avec les règles ci-dessus, la stratégie gagnante est de laisser toujours un multiple de trois allumettes à son adversaire. En effet, si on laisse  $3n$  allumettes à son adversaire, il en restera après son coup  $3n - 1$  (s'il en retire une) ou  $3n - 2$  (s'i en retire deux). On peut alors se débrouiller pour lui en laisser  $3n - 3$  (autre multiple de 3) en en retirant deux dans le premier cas, une dans le second. Ainsi, on peut s'assurer de toujours laisser un multiple de trois allumettes à son adversaire. Quand il ne lui en reste plus que trois, il nous en laissera une ou deux après son coup et on pourra prendre la dernière.

On peut remarquer que la machine apprend, dans le cadre qui lui est donné sans connaître le concept de multiple de 3. C'est le renforcement qui lui fait appliquer la bonne stratégie.

Connaissant cette stratégie, il aurait été possible de programmer la machine afin qu'elle suive la stratégie directement en initialisant le jeu ainsi:

- numéro 1 : que des billes jaunes
- numéro 2 : que des billes rouges
- numéro 3 : un mélange de rouges et de jaunes (peu importe ce qu'elle joue, la machine est dans une situation perdante)
- numéro 4 : que des billes jaunes
- numéro 5 : que des billes rouges
- numéro 6 : un mélange de rouges et de jaunes (peu importe ce qu'elle joue, la machine est dans une situation perdante)
- numéro 7 : que des billes jaunes
- numéro 8 : que des billes rouges

C'est un cas, où l'apprentissage automatique n'était pas nécessaire. Par contre, cela permet bien d'illustrer son principe.

Si l'on initialise le programme simulant le jeu avec ce contenu, la machine gagnerait tout de suite à chaque partie.

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