

## Des réseaux de neurones artificiels apprennent la natation

*Des techniques inspirées par l'étude du fonctionnement du cerveau humain permettent aux robots d'apprendre à coordonner leurs mouvements.*

Nager demanderait-il plus de capacités intellectuelles que jouer aux échecs? Cette question peut paraître étonnante, mais il y a de quoi s'interroger : Alors que cela fait déjà cinq ans que le champion de monde d'échecs s'est fait battre par une machine, la construction de systèmes artificiels dotés de capacités motrices comparables à celles des humains représente encore un grand défi. Au-delà de la difficulté de construire des mécanismes ayant les qualités des muscles et du squelette des animaux, le problème de l'intelligence motrice se pose : étant donné un certain nombre de capteurs et de moteurs, comment construire un «cerveau» qui fasse le lien entre eux? Comment organiser la coordination des commandes envoyées aux moteurs de manière à marcher ou nager de la façon la plus efficace possible? Les mouvements que nous effectuons sans effort intellectuel conscient, avec vitesse et précision, mettent en jeu des mécanismes extrêmement subtils dont les performances restent difficiles à reproduire.

Pour surmonter cette difficulté, des systèmes automatiques imitant certains aspects du fonctionnement du cerveau humain ont été mis au point. Inspirés par des travaux en psychologie et en neurophysiologie, l'apprentissage par renforcement et les réseaux de neurones artificiels sont des méthodes qui ont obtenu de nombreux succès pour la résolution de problèmes de sélection de l'action.

### L'apprentissage par renforcement

Le principe de base de cette forme d'apprentissage est simple : si, dans une situation donnée, une action est accompagnée ou suivie d'une sensation agréable (un renforcement positif), alors cette action aura tendance à être préférée si la même situation se présente à nouveau. Inversement, si l'action est accompagnée ou suivie d'une sensation désagréable (un renforcement négatif), alors cette action aura tendance à ne pas être choisie la prochaine fois que la situation se produira. Ce principe élémentaire de l'intelligence humaine, établi par les psychologues, a été repris pour construire des algorithmes d'apprentissage qui peuvent être programmés sur des ordinateurs.

Une des premières applications de cette technique fut la construction de programmes informatiques qui apprennent à jouer au morpion (le jeu dont le but est d'aligner trois X ou trois O dans une grille  $3 \times 3$ ). Ici, le renforcement est +1 en cas de victoire, -1 en cas de défaite et 0 sinon. Un ordinateur peut apprendre à jouer en stockant une table de valeurs qui, pour chaque position possible, indique une estimation de l'espérance de gain final. Initialement, cette table est

remplie avec des valeurs aléatoires. Durant la partie, l'ordinateur sélectionne chaque coup en choisissant de préférence celui qui mène vers la position qui a la plus grande valeur d'après la table. De manière à explorer toutes les positions possibles, il essaie aussi de temps en temps des coups aléatoires. A la fin de chaque partie, selon qu'il s'agit d'une victoire, d'une nulle ou d'une défaite, les valeurs des positions qu'il a rencontrées au cours du jeu sont légèrement modifiées pour se rapprocher de +1, de 0 ou de -1. Au fur et à mesure des parties, la table de valeurs devient de plus en plus exacte et le niveau de jeu augmente.

En théorie, cette méthode pourrait s'appliquer à n'importe quel problème de sélection de l'action. Il suffirait de mémoriser, pour chaque situation possible, une valeur qui indique l'espérance du renforcement futur, et de mettre à jour cette valeur en fonction de l'expérience. Mais pour un problème comme celui du contrôle moteur d'un robot muni de nombreux capteurs, cette méthode ne peut pas être utilisée. En effet, si les capteurs peuvent prendre des valeurs continues, alors le nombre de situations possibles est infini. Même en supposant que les capteurs ne prennent qu'un nombre fini de valeurs, le nombre de situations croît de manière exponentielle avec le nombre de capteurs, ce qui rend le problème impossible à traiter en pratique. En fait, pour peu qu'il se déplace dans un environnement suffisamment complexe, un robot ne rencontrera jamais deux situations parfaitement identiques. Un algorithme basé sur une table comme celui proposé pour le morpion ne pourra donc pas fonctionner.

De manière à pouvoir prendre une décision dans une situation inédite, un robot devra être capable de *généraliser* son expérience passée. C'est à cela que vont servir les réseaux de neurones artificiels.

### Les réseaux de neurones artificiels

Le terme «réseaux de neurones artificiels» désigne des circuits composés de petites unités de calcul interconnectées, dont le fonctionnement est inspiré de celui du système nerveux humain. Une grande variété de tels systèmes a été conçue pour diverses applications en intelligence artificielle et en traitement de l'information. Un type particulier de réseau de neurones, le perceptron multi-couches, peut effectuer la généralisation dont on a besoin pour l'apprentissage par renforcement.

Le perceptron multi-couches (Figure 1) se présente sous la forme d'un circuit calculant une sortie en fonction d'un certain nombre de valeurs numériques données en entrée. Il peut ainsi jouer un rôle similaire à celui de la table de valeurs utilisée pour

le morpion : l'entrée du réseau est la situation, et la sortie son évaluation. Durant l'apprentissage, c'est la force de chaque connexion qui est modifiée pour que la fonction calculée se rapproche des récompenses observées (de manière analogue à l'interpolation linéaire où on modifie les paramètres de la droite pour se rapprocher des données : Figures 2 et 3). Le réseau peut ainsi généraliser l'expérience enregistrée : à partir de l'évaluation d'un petit nombre de situations, il peut construire une estimation de toutes les autres situations possibles.

Un des succès les plus impressionnants de cette technique est le joueur de backgammon de Tesauro. Ce programme a atteint un niveau de jeu très supérieur à celui des autres, rivalisant avec les meilleurs joueurs humains. Ce qui est encore plus spectaculaire, c'est que cet apprentissage l'a conduit à inventer un style de jeu original, différent des stratégies répandues à l'époque. Depuis, les joueurs humains se sont mis à imiter le style de l'ordinateur, car il se sont aperçus qu'il est plus efficace ! Ce résultat remarquable a beaucoup contribué à attirer l'attention des chercheurs sur l'apprentissage par renforcement.

Cette méthode peut aussi s'appliquer aux problèmes de contrôle moteur. Jusqu'à présent, son succès se limitait à des mécanismes ayant très peu d'articulations. Dans l'équipe Apprentissage et Cognition du laboratoire Leibniz-IMAG, nous avons perfectionné les techniques existantes pour dépasser cette limite. Cela nous a permis de faire apprendre la natation à des robots simulés comprenant jusqu'à sept segments articulés. La Figure 4 montre un exemple de mouvement obtenu par un nageur à cinq segments. Grâce aux très bonnes capacités de généralisation du réseau de neurones, l'apprentissage par renforcement peut ainsi traiter un problème avec un nombre d'états possibles gigantesque.

Même si ces résultats représentent un petit progrès, on est encore bien loin de la complexité du corps humain, et l'adresse motrice des animaux reste hors de portée des robots. Dans l'avenir, on peut espérer que les collaborations entre biologistes, mathématiciens et roboticiens conduiront à une meilleure compréhension des performances des systèmes naturels, et à la construction de systèmes artificiels encore plus efficaces. Le problème du contrôle moteur occupera encore certainement des générations de chercheurs, car les progrès restant à faire sont immenses, et les pistes à explorer sont très nombreuses.

## Contact

Rémi Coulom (<http://remi.coulom.free.fr/>)

## Pour en savoir plus

*Reinforcement Learning : An Introduction*, par Richard S. Sutton et Andrew G. Barto (consultable en ligne à l'adresse suivante : <http://www-anw.cs.umass.edu/~rich/book/the-book.html>)

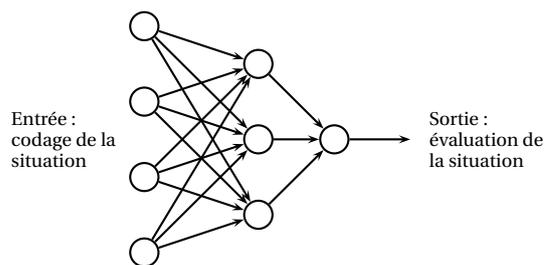


FIG. 1 – Un perceptron multi-couches pour l'apprentissage par renforcement

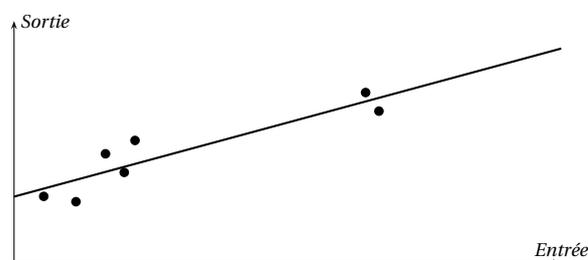


FIG. 2 – L'interpolation linéaire permet de généraliser en ajustant les paramètres de la droite (pente et ordonnée à l'origine) de manière à se rapprocher le plus possible des données.

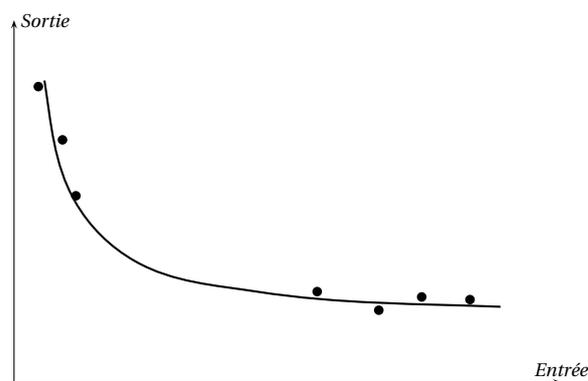


FIG. 3 – De manière similaire à l'interpolation linéaire, en ajustant les forces des connexions pour se rapprocher des données, un perceptron multi-couches peut interpoler des fonctions non-linéaires quelconques.

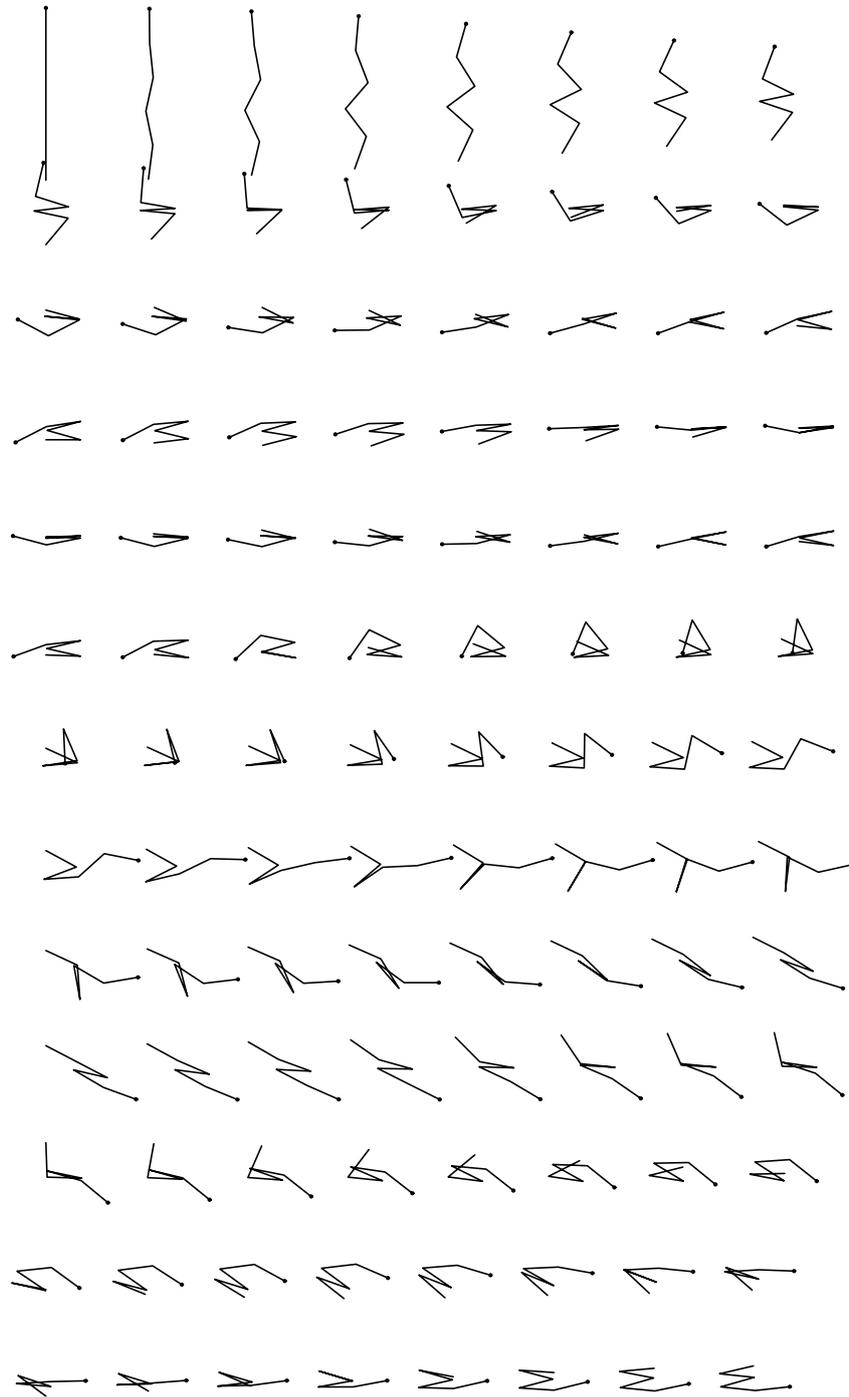


FIG. 4 – Un demi-tour effectué par un nageur à cinq segments. Cette figure se lit de gauche à droite et de haut en bas. Au départ, le nageur est immobile et tous ses segments sont alignés verticalement. Dans les cinq premières lignes, il se déplace vers la droite. Dans les lignes suivantes, il fait demi-tour vers la gauche. (Pour une meilleure visualisation, des démonstrations interactives de ces nageurs peuvent être téléchargées à cette adresse : [http://remi.coulom.free.fr/Thesis/.](http://remi.coulom.free.fr/Thesis/))