

2016

# Apprentissage profond

## pour intelligences artificielles

Après des décennies de trop lents progrès, les recherches en intelligence artificielle ont franchi un pas décisif, qui explique l'effervescence actuelle autour des réseaux neuronaux profonds.

**E**n 2016, un ordinateur a battu Lee Sedol, champion professionnel du jeu de go, grâce à un logiciel conçu par la société Google DeepMind. Cet exploit remarquable est le fruit de grands et récents progrès dans le domaine de l'intelligence artificielle. Mais est-on pour autant proche de l'intelligence humaine? Non, car le logiciel de DeepMind ne sait rien faire d'autre que jouer au go. Comment en est-on arrivé là et comment envisage-t-on de progresser?

L'intelligence se démontre à travers une compétence dans des domaines particuliers, ce qui nécessite de bonnes connaissances dans les domaines en question. Pour rendre les ordinateurs aussi intelligents que les humains, il faut donc faire en sorte que les ordinateurs aient au moins autant de connaissances qu'un humain ordinaire. Plusieurs décennies durant, l'approche dominante dans les recherches sur l'intelligence artificielle a ainsi consisté à fournir directement à la machine des connaissances.

Il s'agissait alors de formaliser nos connaissances, c'est-à-dire de les mettre sous une forme claire et compréhensible par l'ordinateur. Mais beaucoup de nos connaissances sont intuitives et ne nous sont pas consciemment accessibles.

L'intuition est ce qui permet au joueur de go de bien choisir son prochain coup, au bon conducteur d'éviter un accident, au médecin d'allier ses connaissances livresques à la psychologie du patient; et surtout, cela implique presque toutes nos capacités perceptives, que l'on ne sait pas décrire précisément, mais qui sont essentielles à notre compréhension du monde.

Pour l'intelligence artificielle, la solution qui s'est révélée fonctionner à merveille est que l'ordinateur acquière lui-même ses connaissances, en apprenant à partir d'exemples. Le domaine scientifique relatif à cet «apprentissage automatique» a connu plusieurs vagues de progrès, et l'on y a exploré de nombreux algorithmes et formalismes mathématiques. Une bonne partie d'entre eux sont regroupés derrière l'idée des «réseaux de neurones», qui consiste à s'inspirer du cerveau pour effectuer des formes de calcul différentes de celles de l'ordinateur classique.

Dans cette approche, des neurones artificiels (des éléments logiciels ou matériels) reçoivent des signaux provenant d'autres neurones, font un calcul relativement simple pour combiner ces signaux et renvoient le résultat à d'autres neurones (voir l'encadré page 102). Les réseaux neuronaux artificiels sont souvent organisés en couches successives, les résultats calculés par une couche étant envoyés aux couches suivantes.

### L'AUTEUR



**YOSHUA BENGIO**  
professeur  
au département  
d'informatique  
et recherche  
opérationnelle  
de l'université  
de Montréal



Fondé sur de l'apprentissage profond, le programme AlphaGo de la société Google DeepMind a permis à un ordinateur (plutôt qu'un robot) de battre au jeu de go un champion mondial, le Coréen Lee Sedol.

L'apprentissage se fait par une modification du calcul effectué par chaque neurone, selon un algorithme qui ajuste les «poids» des connexions entre neurones (des coefficients dont dépend le calcul fait par chaque neurone).

Au début des années 2000, les recherches sur les réseaux neuronaux n'étaient plus en vogue, faute de résultats suffisamment probants. Or c'est précisément à ce moment qu'une percée a eu lieu, hors de la lumière des projecteurs médiatiques qui inondent aujourd'hui le domaine. Pour la première fois, en 2006, trois équipes, autour de Geoffrey Hinton, à l'université de Toronto, de moi-même, à l'université de Montréal, et de Yann LeCun, à l'université de New York, ont réussi, avec un nouvel algorithme, à entraîner des réseaux neuronaux «profonds» – composés de plus d'une ou deux couches entre l'entrée et la sortie du réseau. Nous avions l'intuition, depuis longtemps, qu'un réseau profond pourrait être plus puissant que ceux explorés jusque-là. Ce qui a été confirmé dans les années suivantes.

Les tentatives précédentes pour entraîner des réseaux profonds n'avaient pas été fructueuses, sans doute à cause d'une difficulté mathématique que l'on ne comprend pas encore totalement. C'est seulement en 2011 que notre équipe a découvert que certaines formes de non-linéarité dans le calcul effectué par les

neurones artificiels facilitaient beaucoup l'entraînement de réseaux profonds. Depuis, les succès et les progrès se sont accélérés, d'abord pour la reconnaissance de la parole – de nos jours, presque tous les systèmes opérationnels de ce type utilisent l'apprentissage profond –, puis pour la vision par ordinateur.

Ces progrès ont été possibles non seulement grâce à de nouveaux algorithmes et architectures de réseaux neuronaux, mais aussi grâce à la puissance de calcul accrue des puces graphiques (les GPU) et à la disponibilité de grandes bases de données étiquetées (qui contiennent des millions d'exemples servant à l'entraînement, pour lesquels on indique à l'ordinateur la bonne réponse à produire, par exemple la catégorie à laquelle appartient l'objet principal figurant dans une image).

L'ordinateur fait de l'apprentissage supervisé: on le nourrit avec des exemples constitués chacun d'un couple (entrée, sortie désirée). Souvent, la sortie désirée doit être fournie par un humain. Ce travail d'étiquetage, fastidieux et coûteux, est aujourd'hui au cœur de la plupart des systèmes industriels fondés sur l'apprentissage profond.

Alors que les réseaux neuronaux des années 1990 étaient essentiellement limités à la reconnaissance des formes, appliquée à une

## BIBLIOGRAPHIE

Y. LeCun et al., **Deep learning**, *Nature*, vol. 521, pp. 436-444, 2015.

D. Bahdanau et al., **Neural machine translation by jointly learning to align and translate**, *ICLR Conference*, 2015.

X. Glorot et al., **Deep sparse rectifier neural networks**, *AISTATS Proceedings, PMLR*, pp. 315-323, 2011.

G. E. Hinton et al., **A fast learning algorithm for deep belief nets**, *Neural Computation*, vol. 18(7), pp. 1527-1554, 2006.

Y. Bengio et al., **Greedy layer-wise training of deep networks**, *NIPS Proceedings*, pp. 153-160, 2006.

entrée représentée par un vecteur de taille fixe, les réseaux neuronaux d'aujourd'hui peuvent traiter des entrées dont la taille est variable – listes, ensembles, arbres, graphes, etc.

Par ailleurs, certains algorithmes et architectures confèrent aux réseaux neuronaux un mécanisme dit d'attention diffuse, qui permet à chaque instant de focaliser le calcul sur un élément (ou un petit nombre d'éléments) de l'entrée qui est plus pertinent. Ces éléments peuvent être des positions dans une image, auquel cas on parle d'attention visuelle, des parties d'un graphe ou de tout autre ensemble d'éléments que l'ordinateur doit traiter.

Le mécanisme d'attention diffuse facilite l'entraînement du réseau neuronal et a été l'élément principal qui nous a permis d'obtenir une percée en traduction automatique, à partir de 2014-2015. On retrouve maintenant un tel système dans Google Translate, avec une qualité de traduction de loin supérieure à ce qui existait auparavant.

L'attention diffuse intervient aussi dans l'exploration moderne de mécanismes appris de raisonnement. Les chercheurs en apprentissage profond considèrent aujourd'hui le raisonnement comme une façon de combiner séquentiellement plusieurs éléments (des faits) afin de parvenir à une conclusion (un fait inféré ou déduit). Le mécanisme de raisonnement (par exemple, une règle d'inférence du type: « Si A et B, alors C ») lui-même n'est pas fixé d'avance, mais est appris par la machine, de façon à pouvoir résoudre les problèmes qui lui sont soumis et dont la solution exige un raisonnement.

## APPRENTISSAGE NON SUPERVISÉ

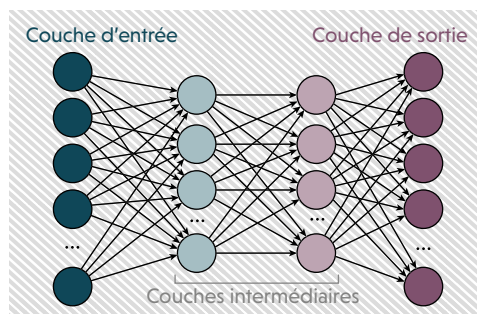
Pour s'approcher de l'intelligence humaine, il faudra découvrir des mécanismes d'apprentissage grâce auxquels la machine pourra réellement comprendre les rouages profonds de son environnement. Or il est impossible de compter sur les humains pour fournir une interprétation sémantique de toutes les données utilisées pour entraîner la machine. C'est pourquoi les chercheurs s'éloignent du cadre classique de l'apprentissage supervisé et se mettent à combiner apprentissage non supervisé et apprentissage « par renforcement ».

Dans l'apprentissage profond non supervisé, on fournit à la machine un ensemble de données ou d'exemples sans lui spécifier le résultat qu'elle doit assigner à chacun d'eux: l'algorithme doit permettre au système de classer de lui-même ces données (par exemple, à partir de nombreuses images de véhicules présentées lors de l'apprentissage, la machine pourra déterminer, sur une nouvelle image, si l'objet représenté est un véhicule ou quelque chose d'une autre nature).

Les chercheurs s'intéressent particulièrement à l'apprentissage profond non supervisé et « génératif », où l'ordinateur peut créer

## RÉSEAUX DE NEURONES ARTIFICIELS

**L**es réseaux de neurones artificiels sont des systèmes, étudiés dès la fin des années 1940, qui s'inspirent des propriétés des neurones biologiques et de leurs connexions.



aléatoirement de nouvelles données intéressantes, par exemple synthétiser de nouvelles images, de nouveaux sons, de nouvelles phrases. La tâche d'un tel ordinateur-artiste n'est pas aisée: il ne suffit pas de produire des pixels au hasard, ces pixels doivent être organisés d'une façon qui soit plausible, qui puisse, par exemple, correspondre à une image naturelle.

Quant à la démarche dite d'apprentissage par renforcement, elle consiste à apprendre à produire une séquence d'actions qui maximisent un ou plusieurs objectifs ou récompenses futures. C'est ainsi que, par exemple, l'enfant apprend une physique intuitive sans l'aide de ses parents, en jouant, en observant et en interagissant avec son environnement.

L'apprentissage non supervisé et par renforcement permettra peut-être à une machine de bien se représenter le monde qui l'entoure (c'est-à-dire de le comprendre) et ainsi de mieux planifier ses actions. Il est difficile d'évaluer le chemin qui reste à parcourir pour atteindre un tel niveau d'intelligence artificielle. Mais ce qui est sûr, c'est que de nombreuses applications utiles seront possibles dès demain grâce aux techniques d'apprentissage profond que nous connaissons déjà (d'où l'engouement de nombreuses entreprises pour ce domaine). Et que ces potentialités seront plus énormes encore quand les recherches fondamentales déboucheront sur de l'intelligence artificielle comparable à celle d'un humain. ■

Un neurone artificiel peut être un composant logiciel ou un composant matériel. À partir des signaux numériques qu'il reçoit d'autres neurones, il calcule, selon des règles simples, un résultat qu'il transmet aux neurones suivants. Ce signal est lui-même pondéré par la « force » – on dit aussi le « poids » – de la connexion par laquelle il passe.

Typiquement, un réseau neuronal est constitué d'une couche de neurones d'entrée, d'une couche de sortie et d'une ou plusieurs couches intermédiaires ou « cachées » (jusqu'à plusieurs centaines dans les réseaux profonds). Les neurones d'entrée recueillent les données (par exemple les valeurs des pixels d'une image), ceux de sortie transmettent un résultat (qui code par exemple le type d'objet que le réseau a reconnu dans l'image).

Dans le processus d'apprentissage, où un grand nombre d'exemples sont présentés au réseau, les poids des connexions sont modifiés peu à peu selon un algorithme approprié, afin d'obtenir des performances optimales de reconnaissance.