

à établir des procédures, nommées algorithmes d'apprentissage, qui confèrent à une machine la capacité d'apprendre à partir d'exemples qu'on lui présente.

La science de l'apprentissage automatique fait face à un résultat négatif formel surnommé *No free lunch* (« Rien n'est gratuit ») : si tous les problèmes sont équiprobables, il n'existe pas d'algorithme universel, c'est-à-dire meilleur que tous les autres

Les réseaux de neurones à apprentissage profond s'inspirent de la structure en couches multiples du cortex visuel, la partie du cerveau qui traite les signaux venant des yeux

sur l'ensemble des problèmes. Il faut donc élaborer des algorithmes différents pour s'attaquer à différentes catégories de problèmes (par exemple reconnaître un coucher de soleil ou traduire un texte en ourdou). L'apprentissage automatique repose ainsi sur deux piliers, théorique et expérimental : un problème réel ne satisfait pas toujours les hypothèses théoriques de l'algorithme, et la validation de l'algorithme sur des données réelles est toujours nécessaire.

Or il semble que notre cerveau incorpore des algorithmes généraux qui nous permettent d'apprendre une multitude de tâches auxquelles l'évolution n'avait pas préparé nos ancêtres : jouer aux échecs, construire des ponts ou faire de la recherche en intelligence artificielle.

Ces compétences supplémentaires suggèrent que l'intelligence humaine pourrait encore être source d'inspiration pour créer des machines dotées d'une forme d'intelligence générale. C'est exactement pourquoi les développeurs de réseaux de neurones ont adopté le modèle du cerveau comme guide pour concevoir des systèmes intelligents.

Les principales unités de calcul du cerveau sont des cellules nommées neurones. Un neurone transmet un signal sous la forme d'une impulsion électrique qui se propage jusqu'à la synapse, la zone de contact avec un autre neurone. Des molécules nommées neurotransmetteurs sont libérées, puis réabsorbées par le neurone cible, ce qui permet à ce dernier de prendre le relais. La propension d'un neurone à transmettre un signal vers un autre neurone est nommée force synaptique. À mesure qu'un neurone « apprend », sa

force synaptique augmente, et il a plus de chances d'envoyer des signaux à ses voisins quand il est stimulé par une impulsion électrique.

Les neurosciences ont influencé l'émergence des réseaux de neurones artificiels, où ces éléments sont connectés de façon matérielle ou logicielle. Les premiers programmeurs de cette sous-discipline de l'intelligence artificielle, connue sous le nom de connexionnisme, postulaient que les réseaux neuronaux seraient capables d'apprendre des tâches complexes en modifiant progressivement les connexions entre neurones, de telle façon que les schémas d'activité neuronale coderaient le

contenu des données – livrées sous forme d'images, de sons ou autres. À chaque exemple soumis au réseau, le processus d'apprentissage se poursuivrait en modifiant les forces synaptiques entre les neurones connectés. Ces valeurs convergeraient petit à petit vers la configuration du réseau qui représente le mieux, par exemple, le contenu d'images d'un coucher de soleil.

Du cerveau aux réseaux virtuels

Les réseaux de neurones actuels sont des versions améliorées des travaux pionniers issus du connexionnisme. Cependant, les algorithmes d'apprentissage correspondants requièrent une participation active de l'homme. La plupart d'entre eux utilisent une technique, nommée apprentissage supervisé, où chaque exemple d'apprentissage est accompagné d'une étiquette indiquant ce qui fait le sujet de l'apprentissage (une image de coucher de soleil, par exemple, est associée à l'étiquette « coucher de soleil »).

Dans ce cas, l'objectif de l'algorithme d'apprentissage supervisé est, à partir d'une donnée d'entrée constituée par une photographie, de produire comme sortie le nom de l'objet central de l'image. L'opération mathématique qui transforme une entrée en une sortie est une fonction. Les valeurs numériques qui définissent cette fonction, telles que les forces synaptiques, correspondent à une solution de la tâche d'apprentissage.

Il serait facile pour un système d'apprendre par cœur les réponses correctes

Apprentissage profond et étude de l'ADN

Dans le domaine de la génomique, l'accumulation de données nécessite des algorithmes puissants pour en extraire des informations. En 2015, Brendan Frey, de l'université de Toronto, et ses collègues ont développé DeepBind, un logiciel qui recourt à l'apprentissage profond pour analyser la façon dont des protéines se lient à l'ADN et à l'ARN. Les chercheurs détectent ainsi des mutations cellulaires entraînant des maladies.

sur les exemples connus (il suffit d'une mémoire de capacité suffisante). Mais cela n'a pas grand intérêt. Même si on accumule des millions d'images de coucher de soleil, le nombre d'images possibles d'une telle scène est infini. Ainsi, l'objectif de l'apprentissage est d'être capable de donner de bonnes réponses sur de nouveaux exemples, inconnus du système, en généralisant les exemples connus. Le bon niveau de généralisation dépend du contexte. Ainsi, on peut vouloir reconnaître la notion d'arbre en général, mais on peut aussi vouloir distinguer les chênes des hêtres, ou vouloir reconnaître le hêtre d'un jardin donné...

Une structure faite de couches multiples

Un tel algorithme doit aussi s'appuyer sur certaines hypothèses relatives aux données et à ce que pourrait être une solution possible à un problème donné. Par exemple, le logiciel doit intégrer comme principe que si des données d'entrée d'une fonction particulière sont semblables, leur sortie ne devrait pas être radicalement différente : modifier quelques pixels sur une image de chat ne devrait pas transformer l'animal en chien.

De telles hypothèses faites sur des images sont utilisées par les réseaux de neurones dits convolutifs. Ces programmes sont à l'origine du renouveau de l'intelligence artificielle. Les réseaux neuronaux convolutifs utilisés dans l'apprentissage profond comprennent de nombreuses couches de neurones organisées de telle manière que le logiciel sera robuste vis-à-vis des changements dans l'objet qu'il tente d'analyser. Il sera capable de reconnaître l'élément même s'il a un peu bougé par exemple. Ainsi, un réseau bien entraîné sera capable de reconnaître un visage que les photographies représentent sous des angles variés.

La configuration d'un réseau convolutif s'inspire de la structure en couches multiples du cortex visuel, c'est-à-dire la partie de notre cerveau qui reçoit les signaux des yeux. Ces nombreuses couches de neurones virtuels justifient le qualificatif de « profond » donné à ces réseaux et confèrent à ces derniers une meilleure capacité à appréhender le monde environnant.

L'apprentissage profond est devenu envisageable il y a une dizaine d'années grâce à certaines innovations, alors que



© Google

LA MACHINE GAGNE AU JEU DE GO.

En mars 2016, le logiciel AlphaGo a affronté le joueur coréen, Lee Sedol, considéré comme l'un des meilleurs du monde (ci-dessus au milieu, avec *Sergei Brin, cofondateur de Google, et Demis Hassabis, cofondateur de Google DeepMind qui a développé AlphaGo*). Fondé sur l'apprentissage profond, AlphaGo a battu l'humain quatre manches à une (ci-contre, la seule partie gagnée par Lee Sedol).



l'intérêt pour l'intelligence artificielle était au plus bas. Un organisme canadien financé par le gouvernement et par des fonds privés, l'ICRA (Institut canadien de recherches avancées), a contribué à ranimer la flamme en soutenant un programme dirigé par Geoffrey Hinton, de l'université de Toronto, et dont faisaient partie Yann LeCun de l'université de New York et du centre de recherche de Facebook à Paris, Andrew Ng de l'université de Stanford, Bruno Olshausen de l'université de Californie à Berkeley, moi-même et plusieurs autres. En raison du scepticisme ambiant, il était difficile à cette époque de publier des articles sur le sujet et même de persuader des étudiants de faire leur thèse dans ce domaine. Mais nous étions convaincus qu'il fallait persévérer dans cette voie.

La réticence vis-à-vis de l'intelligence artificielle découlait initialement de l'idée que la création de réseaux neuronaux était sans espoir à cause de la difficulté à optimiser leur performance pour apprendre efficacement.

L'optimisation est une branche des mathématiques qui vise à trouver la meilleure combinaison de paramètres, ici les poids synaptiques des neurones virtuels, pour atteindre un certain objectif. Lorsque la relation qui lie les paramètres et l'objectif

est assez simple – plus précisément lorsque l'objectif est une « fonction convexe » des paramètres – alors on peut améliorer progressivement les paramètres et l'on parle d'optimisation convexe. La procédure d'entraînement est répétée jusqu'à ce que les paramètres s'approchent aussi près que possible des valeurs qui produisent le meilleur résultat. On cherche en fait un minimum global, à savoir le jeu de paramètres qui permet d'atteindre la valeur la plus basse (et la meilleure) de l'écart à l'optimum.

Des obstacles surmontés peu à peu

En général, la situation n'est pas aussi simple, la fonction reliant les paramètres à l'objectif qu'ils atteignent n'étant pas convexe. Or l'optimisation non convexe représente un défi bien plus difficile. De nombreux chercheurs pensaient qu'il n'était pas possible de le relever. L'algorithme d'apprentissage pourrait en effet se retrouver coincé dans un minimum local, d'où il ne pourrait sortir s'il ajuste les paramètres par petites touches.

Or mes collègues et moi avons montré que si le réseau de neurones est assez grand, le problème des minima locaux est fortement