«Comme être humain, apprendre à reconnaître un chien revient à savoir identifier quatre pattes poilues, analyse Maithra Raghu, chercheuse chez Google Brain. Idéalement, nous aimerions que nos réseaux neuronaux procèdent de la même manière.» L'abstraction vient naturellement au cerveau humain. Pas aux réseaux neuronaux. Ils sont faits de blocs de construction appelés «neurones», connectés de diverses manières. Chacun peut représenter un attribut, ou une combinaison d'attributs, que le réseau prend en compte à chaque niveau d'abstraction. Quand ils relient ces neurones, les ingénieurs doivent faire de nombreux choix, en particulier décider de combien de couches de neurones le réseau doit disposer, autrement dit, quelle est sa «profondeur». Imaginez que l'enjeu soit de reconnaître des objets dans des images. L'image entre dans le système à la première couche. À la couche suivante, le réseau peut avoir des neurones qui détectent simplement les contours dans l'image. Puis la suivante combine les courbes en formes et textures, et la dernière traite formes et textures pour conclure sur la nature de l'objet : c'est un mammouth laineux !

Chaque couche combine plusieurs aspects de la précédente. Ainsi, un cercle peut être vu comme une combinaison de courbes, une courbe comme une combinaison de segments. Les ingénieurs doivent aussi décider de la «largeur» de chaque couche, laquelle correspond au nombre de caractéristiques différentes que le réseau prend en compte à chaque niveau d'abstraction. Dans le cas de la reconnaissance d'image, cette largeur peut être le nombre de types de segments, de courbes et de formes considérés.

Quanta magazine

Cet article a d'abord été publié en anglais par ***Quanta Magazine,*** *une publication en ligne indépendante soutenue par la Simons Foundation afin de favoriser la diffusion des sciences : bit.ly/36dMtRh*

Au-delà de la profondeur et de la largeur d'un réseau, il faut aussi choisir comment câbler les neurones dans et entre les couches, et quel poids donner à chaque connexion. Dans ces conditions, une fois l'objectif fixé, comment savoir quelle architecture de réseau neuronal l'atteindra au mieux ? Quelques règles générales se dégagent. Pour les tâches impliquant des images, les ingénieurs utilisent des réseaux neuronaux «convolutifs», qui offrent le même motif de connexions entre les couches, répété encore et encore. Pour le traitement du langage naturel – reconnaissance automatique de la parole ou génération de textes –, les réseaux neuronaux «récurrents» semblent mieux fonctionner : les neurones peuvent y être connectés à des couches non adjacentes.

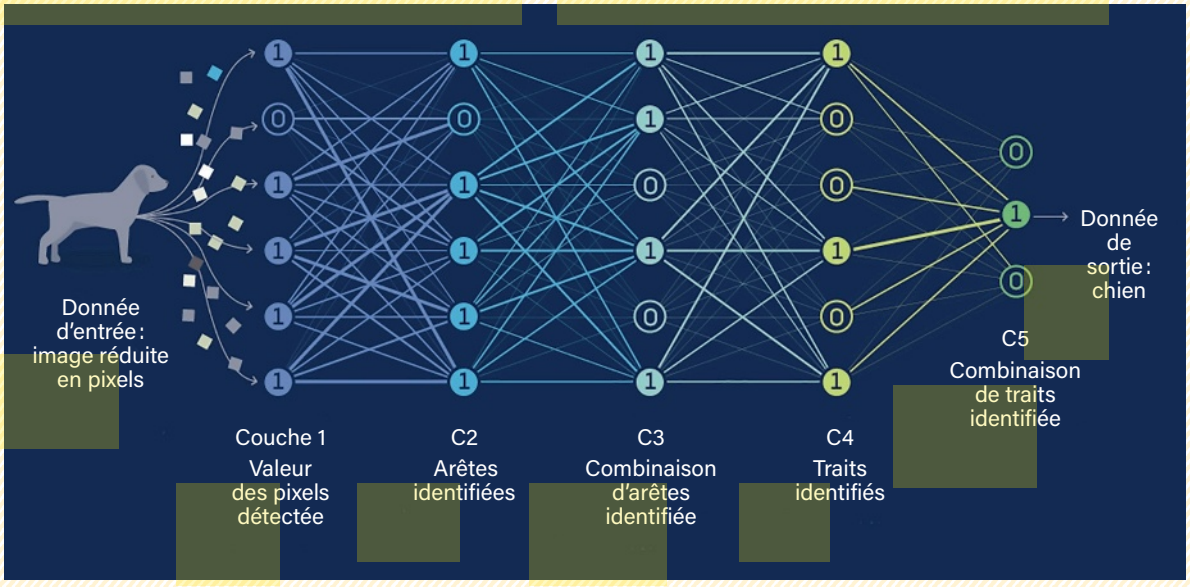
LA RÉUSSITE COMME PRODUIT D'ESSAIS ET D'ERREURS

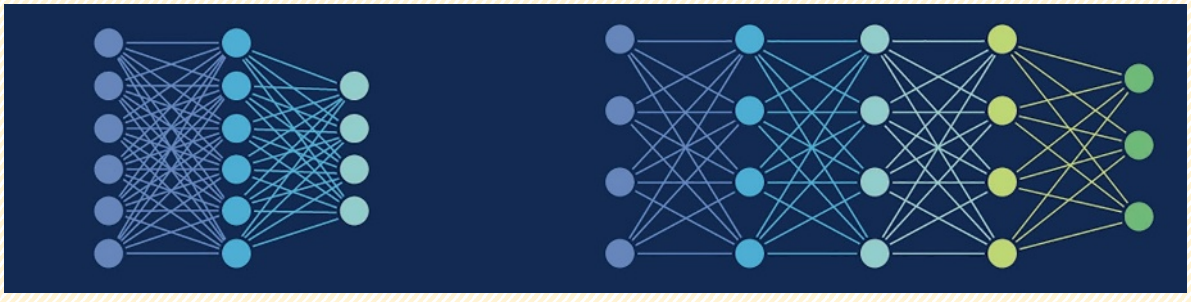
Au-delà de quelques lignes directrices, les ingénieurs n'ont d'autre choix que d'obtenir des preuves expérimentales : ils lancent 1 000 réseaux neuronaux différents et observent simplement lequel fait le boulot. «Leurs choix sont le produit d'essais et d'erreurs, constate Boris Hanin. C'est dur, car il y a un nombre infini de choix possibles et nul ne sait quel est le meilleur.»

Il conviendrait de réduire la quantité d'essais et d'erreurs et de mieux comprendre au préalable ce que l'architecture d'un réseau neuronal peut apporter. Quelques études récentes font avancer dans cette direction. Selon David Rolnick, aujourd'hui professeur de sciences

Comment on conçoit un réseau de neurones

Un réseau neuronal fait circuler des données d'entrée, par exemple des images, à travers des couches de neurones numériques. Chacune d'elles met en évidence de nouvelles caractéristiques de la donnée d'entrée. En fonction de l'architecture mathématique du réseau, c'est-à-dire le nombre de neurones, de couches et leurs interconnexions, les mathématiciens tâchent de définir dans quels types de tâches il excellera. Une fois le réseau alimenté en données, chacun des neurones qui est « excité » (ils sont balisés par un 1 *ci-dessous*) transmet un signal à d'autres neurones de la couche voisine, lesquels s'exciteront à leur tour au-dessus d'un seuil de signaux. Le processus, à base de calculs, extrait des informations abstraites.





Un réseau peu profond, ou expressif, compte peu de couches, mais beaucoup de neurones par couche. Il exige une forte puissance de calcul.

Un réseau profond compte de nombreuses couches, dotées chacune d'assez peu de neurones. Il permet un haut niveau d'abstraction avec une économie de neurones.

informatiques à l'université McGill, « ces travaux visent à rédiger un livre de recettes qui permette de concocter le bon réseau neuronal. Selon ce que vous voulez en tirer, vous choisirez telle ou telle recette ».

L'une des cautions théoriques les plus anciennes concernant l'architecture des réseaux neuronaux a déjà trente ans. En 1989, des informaticiens ont prouvé que si une seule couche informatique est prévue, mais avec un nombre illimité de neurones et autant de connexions entre eux, alors le réseau sera capable d'accomplir n'importe quelle tâche.

Malgré sa formulation carrée, cette affirmation se révèle intuitive et en définitive pas très utile. Elle revient à dire que, si vous pouvez identifier une quantité illimitée de lignes dans une image, vous pouvez distinguer tous les objets en utilisant seulement une couche. Dans le principe, c'est peut-être vrai, mais pour y arriver en pratique… bonne chance !

Aujourd'hui, les chercheurs qualifient d'« expressifs » ces réseaux plats et larges, ce qui signifie qu'ils sont capables en théorie d'embrasser un jeu plus riche de connexions entre données d'entrée (par exemple, une image) et de sortie (ici, une description de cette image). En pratique, ils sont extrêmement difficiles à entraîner, et leur apprendre à produire réellement ces données de sortie est presque impossible. Leur avidité en puissance de calcul dépasse aussi la capacité de nos ordinateurs.

LARGEUR OU PROFONDEUR ?

Plus récemment, les chercheurs ont tenté de cerner jusqu'où ils peuvent pousser les réseaux neuronaux dans la direction opposée : réduire la largeur (moins de neurones par couches) et augmenter la profondeur (davantage de couches). Peut-être suffit-il de repérer 100 lignes différentes, du moment qu'il y a assez de connexions pour transformer ces 100 lignes en 50 courbes, que vous pouvez combiner en 10 formes différentes, lesquelles vous donnent tous les blocs élémentaires nécessaires pour reconnaître la plupart des objets.

En 2018, David Rolnick et Max Tegmark, alors tous deux à l'institut de technologie du Massachusetts (MIT), ont prouvé que plus de profondeur et moins de largeur permettent d'accomplir des fonctions équivalentes, avec une réduction exponentielle du nombre de

> L'abstraction vient naturellement au cerveau humain. Pas aux réseaux neuronaux

neurones. Si vous avez 100 variables d'entrée, on peut obtenir la même fiabilité en utilisant 2^{100} neurones dans une couche ou avec simplement 2^{10} neurones répartis sur deux couches. Combiner de petits éléments à de plus grands niveaux d'abstraction offre plus de potentiel que saisir tous les niveaux d'abstraction d'un seul coup.

« La notion de profondeur dans un réseau neuronal traduit l'idée que vous pouvez réduire n'importe quoi de compliqué en une série de choses plus simples, explique David Rolnick. Comme sur une chaîne de montage. »

David Rolnick et Max Tegmark ont prouvé l'utilité de la profondeur en demandant à des réseaux neuronaux de mener à bien une tâche simple : multiplier des fonctions polynomiales. Il s'agit d'équations dont certaines variables sont élevées à une puissance entière, par exemple $y = x^3 + 1$. Ils ont entraîné les réseaux en leur montrant des exemples d'équations et leur produit. Puis ils leur ont soumis des équations inédites. Les réseaux neuronaux les plus profonds ont appris cette tâche avec beaucoup moins de neurones que ceux qui l'étaient moins.

Bien qu'une multiplication ne soit pas une opération renversante, David Rolnick juge que l'étude marque un point important : « Si un réseau peu profond est incapable de multiplier, nous ne devrions lui accorder aucune confiance. »

D'autres chercheurs évaluent la largeur minimale nécessaire. Toujours en 2018, le mathématicien Jesse Johnson, qui fut chercheur à l'université d'État de l'Oklahoma avant d'être recruté par Sanofi, a prouvé que, à partir