

Deep learning : comprendre les réseaux de neurones artificiels (artificial neural networks)

Le deep learning est de plus en plus développé à notre ère de développement technologique. Il est de plus en plus question de réseaux de neurones artificiels capables de performances surpassant de beaucoup celles d'un cerveau humain. Voici ce qu'il faut comprendre au sujet du système informatique des réseaux de neurones artificiels (artificial neural networks).

Le deep learning

Le deep learning n'est pas seulement une question de technologie, mais aussi de vision globale de l'être humain. L'un des premiers objectifs du deep learning est de **reproduire les capacités de l'être humain** et même d'en créer une version améliorée.

Par deep learning, on entend **un mode d'apprentissage automatique** géré par un réseau de neurones artificiels. Le réseau est composé de plusieurs couches dotées de fonctions précises.

Le deep learning tend donc vers la simulation du cerveau humain en activant ces couches neuronales qui interagissent et fournissent un apprentissage progressif à partir de grands volumes de données. Comme les neurones du cerveau, les **neurones artificiels** arrivent donc à communiquer pour favoriser l'apprentissage et l'assimilation de divers éléments d'information.

Le réseau de neurones artificiels : c'est quoi exactement ?

Le réseau de neurones artificiels **copie le cerveau humain** pour favoriser l'apprentissage. Il s'agit donc d'un système qui se base sur le fonctionnement du cerveau humain pour **l'adapter à des ordinateurs** équipés de fonctions d'intelligence artificielle.

Grâce au réseau de neurones artificiels, l'ordinateur arrive à **résoudre des problèmes de manière autonome**. Le réseau **améliore aussi les capacités de l'ordinateur**.

L'origine des neurones artificiels

Le concept de réseaux de neurones artificiels a vu le jour en 1943 et a été inventé par **Warren McCulloch**, neurophysicien, et **Walter Pitts**, mathématicien. Tous deux étaient chercheurs à l'Université de Chicago. Leur théorie a été présentée dans le Journal Brain Theory. Dans cet article, les deux chercheurs expliquent que l'unité de base de l'activation du cerveau **est l'activation des neurones**.

En développant cette théorie, les chercheurs ont commencé à entrevoir la possibilité de recréer le fonctionnement des neurones. L'année 1957 marque une autre étape dans la progression de la recherche en intelligence artificielle avec l'invention du **perceptron**. Il s'agit d'un algorithme d'apprentissage de classifieurs binaires qui représente le plus ancien algorithme de **Machine Learning**. À partir de cet algorithme, les machines finiront par reconnaître des objets sur des images.

Aujourd'hui, le terme *perceptron* désigne aussi un réseau à propagation avant à couche unique.

Les ordinateurs de l'époque n'étaient **pas assez puissants** pour traiter le volume de données nécessaire pour faire **fonctionner le réseau de neurones artificiels**. La façon d'y arriver était cependant connue et ce n'était

qu'une question de temps avant que les chercheurs ne réussissent à mettre en branle le processus d'apprentissage. Faute de moyens, la recherche en matière de **deep learning** a stagné pendant plusieurs années.

Avec les années 2010 et l'émergence du **big data**, la puissance nécessaire pour créer des réseaux de neurones sophistiqués a été obtenue. Il est même devenu possible de surpasser l'être humain dans le domaine de la reconnaissance d'image. Les limites des réseaux de neurones sont maintenant sans cesse repoussées, ce qui permet d'envisager des avancées spectaculaires dans les années futures.

Le fonctionnement du réseau de neurones artificiels

Le réseau de neurones artificiels est basé sur plusieurs processeurs fonctionnant en parallèle. Ces processeurs sont organisés en tiers. Le premier tiers a pour fonction de recevoir les entrées de données brutes. Chacun des tiers reçoit ensuite les sorties d'informations transmises par le tiers précédent. Le dernier tiers est chargé de produire les résultats du système. Plus le problème est complexe, plus il faut de couches pour le traiter.

Chaque neurone a une valeur particulière qui détermine quelle information peut être transmise au système. La fonction d'activation permet quant à elle de calculer la valeur de sortie de chaque neurone. C'est ce calcul qui détermine combien de neurones doivent être activés pour résoudre le problème. Un algorithme est ensuite créé. Il fait correspondre un résultat à chacune des entrées.

L'algorithme permet à l'ordinateur d'apprendre à partir des nouvelles informations qu'il reçoit. Le réseau de neurones dont est doté l'ordinateur fait en sorte qu'il analyse des exemples de manière à devenir capable d'effectuer une tâche. Ces exemples sont étiquetés. C'est ce processus qui a rendu les ordinateurs capables de reconnaître des objets sur des images, parfois mieux que le cerveau humain lui-même.

Comme pour le cerveau humain, les réseaux de neurones artificiels ne peuvent pas être programmés directement, mais doivent apprendre en étudiant et en analysant des exemples. Il existe trois méthodes d'apprentissage, soit :

- **l'apprentissage supervisé .**
- **l'apprentissage non supervisé .**
- **l'apprentissage renforcé.**

L'apprentissage supervisé

En ce qui concerne l'apprentissage supervisé, l'algorithme s'entraîne à partir de **données étiquetées**. Pour effectuer la tâche, il se modifie jusqu'à arriver à traiter le jeu de données pour obtenir le résultat attendu. Un résultat concret doit être défini pour chaque option d'entrée. L'apprentissage supervisé permet d'apporter des modifications au système afin d'optimiser le fonctionnement de l'algorithme.

L'apprentissage non supervisé

En apprentissage non supervisé, le réseau de neurones doit **analyser un ensemble de données qui ne sont pas étiquetées**. Une fonction spécifique lui indique à quel degré il s'éloigne ou s'approche du résultat attendu. Le réseau s'adapte ensuite. Le résultat de la tâche n'est donc pas déterminé à l'avance, mais le système pose lui-même son diagnostic à partir des informations obtenues. Le système s'appuie entre autres sur la théorie de **la résonance adaptative**.

L'apprentissage renforcé

Quant à l'apprentissage renforcé, il s'agit d'une méthode par laquelle on procède par **renforcements et sanctions** selon que les résultats sont positifs ou négatifs. Comme le cerveau humain qui apprend par essais et erreurs, le réseau de neurones apprend progressivement à mesure qu'il traite les données qu'on lui soumet.

Les différents types de réseaux de neurones

Les **différents types de réseaux de neurones** sont généralement définis selon le nombre d'épaisseurs nécessaire entre l'entrée des données et le résultat final. On détermine aussi le type de réseau à partir du nombre de nœuds cachés dans chaque modèle. Le nombre d'entrées et de sorties de chaque nœud est aussi pris en compte.

Le **réseau de neurones de base est appelé *feed-forward***. Les informations à l'intérieur de ce type de réseau passent directement de l'entrée aux nœuds de traitement. De là, ils sont acheminés directement aux sorties.

Quant aux **réseaux de neurones récurrents**, d'un degré de complexité plus élevé, ils sont capables de sauvegarder les résultats obtenus après le passage de l'information par les nœuds de traitement. Le modèle est peu à peu nourri et façonné à partir des résultats sauvegardés. L'information peut circuler dans une boucle de rétroaction et revenir vers une couche précédente. De cette façon, une mémoire est constituée à l'intérieur du système.

Pour ce qui est des **réseaux de neurones convolutifs**, ils permettent de détecter des motifs simples à l'intérieur d'une image pour en arriver à en identifier le contenu en procédant par recoupements. Leur utilisation est de plus en plus répandue dans des domaines variés, comme la reconnaissance faciale et la numérisation de texte. Ils comportent un minimum de cinq couches. Le résultat obtenu passe d'une couche à l'autre.

Les spécialistes parlent aussi de **réseaux de neurones à propagation avant**. Ces réseaux transmettent l'information dans un seul sens. Les réseaux peuvent être dotés d'une couche unique ou comporter plusieurs couches cachées.

