

Machine learning

Pour faire simple, on peut envisager l'IA comme une série de concepts imbriqués ou dérivés apparus sur une période de plus de 70 ans :

1950 Début des programmes AI.

1980 Les programmes IA apprennent avec les résultats (début du Machine Learning)

2010 Les programmes IA copie le raisonnement humain (début du Deep learning)

2020 Les programmes IA génère des phrases ou des images. (début de IA générative).

Le machine learning, consiste à créer des modèles et à générer des prédictions ou à prendre des décisions sur la base de données. Il englobe un large éventail de techniques qui permettent aux ordinateurs d'apprendre et de faire des choix basés sur les données sans être explicitement programmés pour des tâches spécifiques.

Il existe de nombreuses techniques ou algorithmes de machine learning, notamment la régression linéaire, la régression logistique, les decision trees, les forêts aléatoires, les machines à vecteurs de support (SVM), les k plus proches voisins (KNN), le clustering et autres méthodes. Chacune de ces approches est adaptée à différents types de problèmes et de données.

Parmi les techniques de machine learning utilisées, l'une des plus connues est le réseau de neurones. Les réseaux de neurones s'inspirent de la structure et du fonctionnement du cerveau humain. Ils sont constitués de couches de nœuds interconnectés (analogues aux neurones) qui fonctionnent ensemble pour traiter et analyser des données complexes. Les réseaux de neurones sont bien adaptés aux tâches qui impliquent l'identification des schémas et des relations complexes dans de grandes quantités de données.

La forme la plus simple de machine learning est l'apprentissage supervisé, qui consiste à utiliser des ensembles de données étiquetées pour entraîner des algorithmes afin de classer les données ou de prédire les résultats avec précision.

Dans le cadre de l'apprentissage supervisé, des humains associent chaque exemple d'entraînement à une étiquette de sortie. L'objectif est que le modèle apprenne le mappage entre les entrées et les sorties des données d'entraînement, afin de pouvoir prédire les étiquettes de nouvelles données inédites.

Deep Learning

Le deep learning est une sous-catégorie du machine learning qui utilise des neural networks multicouches, appelés réseaux neuronaux profonds, qui simulent plus étroitement le pouvoir de décision complexe du cerveau humain.

Les réseaux neuronaux profonds comprennent une couche d'entrée, au moins trois, mais généralement des centaines de couches cachées, et une couche de sortie, contrairement aux neural networks utilisés dans les modèles classiques de machine learning qui ne comportent généralement qu'une ou deux couches cachées.

Ces couches multiples permettent un apprentissage non supervisé : elles peuvent automatiser l'extraction de caractéristiques à partir de vastes ensembles de données non étiquetées et non structurées, et faire leurs propres prédictions à partir des représentations de ces données.

L'apprentissage profond ne nécessitant pas d'intervention humaine, il permet de déployer le machine learning à grande échelle. Il est bien adapté au traitement automatique du langage naturel (NLP), à la vision par ordinateur et à d'autres tâches qui impliquent l'identification rapide et précise de modèles et de relations complexes dans de grandes quantités de données. La plupart des applications d'IA que nous utilisons au quotidien sont alimentées par une forme d'apprentissage profond.

Schéma montrant comment les données sont traitées dans un neural network profond Dans un réseau neuronal profond, plusieurs couches de nœuds peuvent extraire le sens et les relations de grands volumes de données non structurées et non étiquetées.

L'apprentissage profond permet également :

L'apprentissage semi-supervisé, qui combine l'apprentissage supervisé et non supervisé en utilisant à la fois des données étiquetées et non étiquetées pour entraîner des modèles d'IA à des tâches de classification et de régression.

L'apprentissage autosupervisé, qui génère des étiquettes implicites à partir de données non structurées, plutôt que de s'appuyer sur des ensembles de données étiquetées pour les signaux de supervision.

L'apprentissage par renforcement, qui apprend par essai-erreur et récompense plutôt qu'en extrayant des informations à partir de modèles cachés.

L'apprentissage par transfert, dans lequel les connaissances acquises dans le cadre d'une tâche ou d'un ensemble de données sont utilisées pour améliorer les performances du modèle dans une autre tâche connexe ou un ensemble de données différent.

(Pour IA Générative voir le document du même nom.)