

QU'EST-CE QUE L'INTELLIGENCE ARTIFICIELLE ?

Livret de vulgarisation

**Mission Villani sur
l'intelligence artificielle**

Mars 2018

Définir l'intelligence artificielle (IA) n'est pas chose facile. Le champ est si vaste qu'il est impossible de la restreindre à un domaine de recherche spécifique ; c'est plutôt un programme multidisciplinaire. Si son ambition initiale était d'imiter les processus cognitifs de l'être humain, ses objectifs actuels visent plutôt à mettre au point des automates qui résolvent certains problèmes bien mieux que les humains, par tous les moyens disponibles. Ainsi l'IA vient au carrefour de plusieurs disciplines : informatique, mathématique (logique, optimisation, analyse, probabilités, algèbre linéaire), sciences cognitives... sans oublier les connaissances spécialisées des domaines auxquelles on souhaite l'appliquer. Et les algorithmes qui la sous-tendent reposent sur des approches tout aussi variées : analyse sémantique, représentation symbolique, apprentissage statistique ou exploratoire, réseaux de neurones, etc.

Définitions de l'IA

Imiter les fonctions humaines

Marvin Lee Minsky, l'un des précurseurs de la discipline définit l'intelligence artificielle comme « la construction de programmes informatiques qui s'adonnent à des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisante par des êtres humains car elles demandent des processus mentaux de haut niveau tels que : l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique ».

En d'autres termes, une intelligence artificielle est avant tout un programme informatique visant à effectuer, au moins aussi bien que des humains, des tâches nécessitant un certain niveau d'intelligence. L'horizon à atteindre concerne donc potentiellement

l'ensemble des champs de l'activité humaine : déplacement, apprentissage, raisonnement, socialisation, créativité, etc.

Les promesses non tenues des débuts de l'IA ont amené à distinguer d'une part les machines qui non seulement mettraient en œuvre des raisonnements semblables aux raisonnements humains, mais auraient également une réelle conscience d'elles-mêmes : c'est ce qu'on appelle **l'intelligence artificielle forte** ; d'autre part les machines qui rendent de nombreux services aux humains en simulant l'intelligence humaine : c'est **l'intelligence artificielle faible**.

L'objectif de la recherche en IA

L'intelligence artificielle forte suscite de nombreux débats autour de l'apparition possible d'une singularité, où la machine, supérieure à l'être humain et consciente de cette supériorité, le supplanterait dans la société. À ce jour, nous en sommes très loin et la majorité des chercheurs en IA pensent même que c'est impossible.

L'intelligence artificielle faible va mettre en œuvre toutes les technologies à sa disposition pour essayer de rendre le service attendu par l'utilisateur. L'intelligence artificielle a, à l'origine, voulu simuler l'activité du cerveau avec l'hypothèse que nous raisonnions avec des règles d'inférence (approche logique de l'IA) ou plus tard, à partir des années 80 avec des neurones formels puis des réseaux de neurones (à l'origine de l'apprentissage profond que nous présentons plus loin). Les progrès dans les algorithmes, les logiques formelles, les puissances de calcul, la standardisation des langages informatiques d'un côté, les sciences du vivant et les sciences cognitives de l'autre, ont fait faire de gros progrès à l'IA qui a pu, dans chacun des domaines

Qu'est-ce que l'IA ?

de recherche qui sont les siens (représentation des connaissances, traitement automatique des langues, robotique, apprentissage, planification et recherche heuristique, modélisation cognitive, etc.), résoudre des problèmes de plus en plus complexes et créer des systèmes qui interagissent de façon fluide et efficace avec les êtres humains.

Dans ce contexte d'IA faible, les systèmes sont de plus nombreux et spécialisés. Ainsi, concrètement, les algorithmes qui créent nos playlists favorites, sélectionnent les meilleurs résultats de recherche ou taguent nos amis sur les médias sociaux ne savent rien faire d'autre ! Plus complexe, un système de traitement automatique du langage naturel chargé de traiter des dépêches d'agence de presse saura ordonner dans le temps divers événements, sans pour autant comprendre le contenu de ces dépêches. L'adaptation de tels systèmes à un autre type de situation demanderait un nouveau cycle de mise au point.

Faire de l'IA : comment s'y prendre ?

IA symbolique : utiliser la logique

L'IA symbolique désigne schématiquement une catégorie d'approches fondée sur des modèles, que ce soit des modèles d'interaction pour la robotique ou les systèmes multi-agents, des modèles syntaxiques et linguistiques pour le traitement automatique des langues, ou des ontologies (ensembles de liens entre mots et concepts) pour la représentation des connaissances. Ces modèles peuvent ensuite, entre autres, être utilisés par des systèmes de raisonnement logique pour produire des nouveaux faits à partir des caractéristiques fournies en entrée. Un des exemples les plus emblématiques

de technique d'IA symbolique est celui des systèmes experts des années 80, aptes à produire un raisonnement à partir de faits et de règles connus. Par exemple, des règles définies par un médecin étaient utilisées telles quelles pour produire automatiquement des diagnostics. Depuis, ce paradigme a largement évolué sur 3 points :

1) Les modèles sont des classifications a priori des objets, qui sont rangés dans des catégories fixes suivant des caractéristiques définies. Ce sont maintenant des ontologies qui, en médecine peuvent contenir des milliers, voire des dizaines de milliers d'objets – i.e. de concepts ;

2) la médecine s'est fortement protocolisée cette dernière décennie et les spécialités médicales élaborent des guides de bonnes pratiques qui peuvent servir de base à une modélisation en IA ;

3) enfin, on a changé de mode d'appréhension du raisonnement formel et du raisonnement médical humain : on est passé du système expert qui allait remplacer le médecin à un système à base de connaissances qui prend en compte les guides de bonnes pratiques et met à la disposition du médecin un système qui va « collaborer » avec lui dans le but de l'aider dans sa tâche.

L'apprentissage automatique

À l'inverse, les techniques d'apprentissage se sont développées en tentant de modéliser les processus cognitifs non à partir d'un ensemble de règles formelles logiques, mais à partir de l'analyse d'expériences passées.

Toutes les approches d'apprentissage automatique comportent deux phases : la première est celle de l'apprentissage à proprement parler et consiste à choisir un modèle (ex : réseau de neurones) puis ajuster ses paramètres à

partir de données en entrée — par exemple des photos de chat et de chien, pour un modèle de reconnaissance visuelle. La deuxième phase est celle de l'inférence. À partir des paramètres qui ont été appris, l'algorithme effectue la tâche qui lui a été fixée — par exemple distinguer les photos de chat des photos de chien.

Parmi les approches de l'apprentissage automatique, on peut distinguer différentes techniques qui peuvent correspondre à différents problèmes.

Apprentissage supervisé

Dans l'apprentissage supervisé, le problème est de distinguer des catégories d'objets dont on dispose des données étiquetées suivant ces catégories. Par exemple des millions de photos où les chats et les chiens sont identifiés comme tels. La phase d'apprentissage ajuste progressivement les paramètres du modèle pour que l'erreur soit la plus faible possible sur les exemples connus. Dans la phase d'inférence, la machine catégorise des données inconnues — elle reconnaît par exemple des chiens sur des nouvelles photos.

Apprentissage non supervisé

Dans l'apprentissage non supervisé, au contraire, on dispose de données non étiquetées. Lors de la phase d'apprentissage, le but est d'établir des catégories. La phase d'inférence est identique à celle de l'apprentissage supervisé. Par exemple, cette technique est particulièrement utile pour déterminer des profils de comportement sur lesquels on n'a aucun a priori sur les catégories, à partir de multiples traces qu'on laisse quotidiennement sur Internet.

Ces approches sont aujourd'hui au centre des travaux de recherche puisqu'elles permettraient de limiter la nécessité de jeux de données annotées.

Apprentissage par renforcement

L'apprentissage par renforcement se distingue des autres modes d'apprentissage dans la mesure où il ne fait pas appel à une base d'exemples constituée en entrée. Son apprentissage s'effectue via des interactions avec l'environnement : la machine effectue une action, l'état de l'environnement change en fonction de celle-ci et la machine se voit attribuer une récompense en fonction du résultat obtenu.

La récompense permet à la machine de découvrir progressivement les meilleures actions à effectuer vis-à-vis du résultat qu'elle cherche à obtenir. Un exemple récent de ce type d'apprentissage a été particulièrement médiatisé avec AlphaGo : l'action est un coup, l'environnement est le plateau de jeu avec les pions disposés dessus, et la récompense est le fait de gagner ou non à la fin de la partie. L'apprentissage par renforcement trouve également des applications en matière de robotique.

L'apprentissage profond

L'apprentissage profond (*deep learning* en anglais) est un sous-domaine de l'apprentissage automatique qui a connu un véritable renouveau depuis 2006. Ces modèles sont construits à partir d'éléments de calcul de base appelés neurones, disposés en général en couches successives. Ces neurones sont reliés entre eux par des connexions dont l'importance est ajustée durant la phase d'apprentissage.

Chaque neurone transforme ses entrées en une sortie en appliquant une fonction de transfert à l'ensemble des informations qui lui parviennent par ces connexions. Les réseaux de

Qu'est-ce que l'IA ?

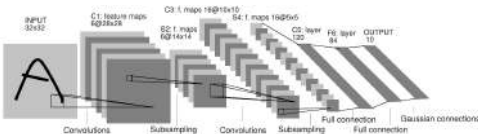
neurons sont des systèmes schématiquement inspirés du fonctionnement du cerveau. Toutefois, il ne s'agit que d'une analogie grossière.

Après apprentissage, on constate que le réseau de neurones a décomposé les données d'entrée en différentes couches d'abstraction hiérarchisées. Par exemple, pour identifier des visages sur une photo, la première couche identifie uniquement les formes élémentaires constituant les visages (lignes, cercles, etc.), la couche suivante les assemble en gros traits du visage (nez, lèvres, front etc..) et la description est progressivement affinée (taille du nez, largeur des narines etc..) jusqu'à recomposer complètement les visages.

Les réseaux de neurones sont particulièrement efficaces dans le traitement de données assimilables à des signaux : son, image, langage, vidéos car ils sont capables d'en capturer les structures complexes sur les plans spatial et temporel.

On peut noter particulièrement deux types de modèles que sont les réseaux de neurones convolutionnels (standard dans le domaine de l'image) et les réseaux de neurones récurrents dans lesquels la notion de couche disparaît. On les utilise surtout pour les signaux temporels et le langage.

Réseau convolutionnel LeNet5 (LeCun, Bottou, Bengio & Haffner, 98)



Applications

Finissons avec quelques applications concrètes.

Reconnaissance d'images et de vidéos

Interpréter une image — reconnaître une personne ou un objet et l'environnement avoisinant — est une chose relativement facile pour un humain. Chaque jour, notre cerveau traite sans grand effort des informations visuelles complexes : une photo de famille, une voiture, un paysage. Mais pour un ordinateur, c'est une tâche très difficile.

Pourtant l'enjeu est de taille, car le développement de la voiture autonome (pour la perception de son environnement), l'annotation automatique d'images, l'amélioration des systèmes d'identification, la détection de pathologies à partir de dispositifs d'imagerie médicale sont tous dépendants des progrès en matière de reconnaissance d'images et de vidéos.

Comment les médias sociaux reconnaissent-ils les visages des personnes présentes sur une photo ?

Ils ont recours à des **réseaux de neurones profonds convolutionnels**. Selon le même principe de l'apprentissage supervisé des images de chiens et chats décrits plus haut, le système apprend à distinguer les visages des amis de l'utilisateur pour lesquels il dispose d'images étiquetées. Lorsqu'une photo est téléchargée sur la plateforme, le système n'a plus qu'à catégoriser les nouveaux visages présents sur celle-ci à partir de sa base de données.

Traduction

Pour construire une application de traduction, on commence par construire une grande base de données composée de textes traduits par des traducteurs humains qui sert de modèle. Il s'agit souvent de livres, de documents officiels produits par des organisations internationales (Nations Unies, Commission européenne) et de sites web qui font autorité. On parle de millions de textes...

Ensuite, un réseau de neurones parcourt ces textes à la recherche de régularités (*patterns*) statistiques. À partir de ces régularités, il construit des équivalences entre des séquences de mots, qui lui permettent de traduire un texte inconnu.

Toutefois, les logiciels de traduction disposent de moins de documents traduits pour certaines paires de langues. Par exemple, il y a plus de documents traduits du français à l'espagnol que du danois au roumain. C'est pour ça que la qualité des traductions varie beaucoup d'une langue à une autre.

Recommandations de contenus

Pour recommander des contenus à leurs utilisateurs, les plateformes de vente en ligne ou de streaming, par exemple, utilisent de l'IA, suivant une autre approche que celles de l'apprentissage mentionnées ci-dessus. Les données de départ sont l'ensemble des choix passés de l'ensemble des utilisateurs. Ces données sont utilisées pour créer des catégories fictives d'utilisateurs et de produits, fictives en ce sens qu'elles représentent des moyennes et non de vrais utilisateurs ou produits : en effet, du fait de l'immense diversité des produits et des utilisateurs, il est impossible de définir directement des catégories à partir de leurs choix de consommation. Chaque

utilisateur peut alors être décrit à partir de ces catégories fictives, ce qui permet de calculer une « proximité fictive » entre les utilisateurs de manière plus fiable que simplement en comptant les produits communs qu'ils ont commandés. On peut alors recommander, pour un utilisateur donné, les produits les plus commandés par l'ensemble des utilisateurs qui leur sont proches suivant cette « proximité fictive ».