

[▶ Je consulte l'article en ligne](#)

[Tribune] Les machines n'apprennent pas

NUMÉRIQUE

PUBLIÉ LE 27/01/2018 À 13H00

"Les logiciels ne sont pas capables de comprendre ce qu'ils manipulent et encore moins capables d'apprendre tout seul." Voilà le point de vue que développe dans cette tribune Christophe Tricot, docteur en informatique et manager expert en intelligence artificielle chez le consultant Kynapse.



2017 était l'année de l'intelligence artificielle et plus particulièrement celle de l'apprentissage automatique ("machine learning" en anglais). Mais est-ce que les machines sont capables d'apprendre ? Cette question est légitime car lorsque l'on utilise des termes comme "intelligence" ou "apprentissage", nous convoquons tout un univers lié à notre expérience et par anthropomorphisme nous accordons trop facilement des qualités aux algorithmes qu'ils n'ont pas. L'apprentissage est l'une d'elles.

La notion d'apprentissage n'est pas simple car au-delà de son apparence triviale, il est nécessaire de mieux saisir cette notion pour comprendre pourquoi il n'est pas question d'apprentissage à proprement parler dans le cadre des algorithmes d'apprentissage automatique.

Tout d'abord regardons la notion d'apprentissage au travers de sa définition dans le TLFi : *L'apprentissage est l'activité qui consiste à "acquérir la connaissance d'une chose par l'exercice de l'intelligence, de la mémoire, des mécanismes gestuels appropriés, etc.*

Cette définition laisse entrevoir la mise en œuvre de processus cognitifs comme la mémorisation et le raisonnement. Pour aller plus loin, regardons du côté de l'étymologie dans le Littré : Apprehendere, *de ad, à, et prehendere, prendre, saisir par l'esprit*

Voilà peut-être l'essence de l'apprentissage chez les humains : saisir par l'esprit. En d'autres termes, apprendre c'est acquérir de nouveaux savoirs en manipulant intellectuellement de nouvelles notions.

Si on regarde du côté de l'apprentissage automatique, nous allons voir qu'il ne s'agit en aucun cas de programmes conscients mais de simples programmes exploitant des modèles configurés automatiquement. Tout est question de modèle.

Pas question de manipuler de nouvelles notions

Dans le cadre des machines et plus particulièrement des techniques d'apprentissage automatique, il n'est pas question de programmes capables de manipuler de nouvelles notions (à l'instar d'un apprenant humain) mais davantage d'algorithmes configurés automatiquement.



L'apprentissage automatique consiste à configurer des algorithmes (comme les réseaux de neurones) en choisissant des poids pour des paramètres

Toutes ces techniques d'apprentissage automatique reposent sur le même principe : configurer automatiquement des modèles mathématiques sur la base de données fournies (on parle de données d'apprentissage). Ces techniques construisent donc les modèles empiriquement.

Par exemple, si je souhaite créer un programme capable de reconnaître la présence d'un chat dans une photo il me faudra construire (ou "entraîner") un modèle de reconnaissance avec d'une part un corpus d'apprentissage constitué de millions de photos dont une grande partie contiennent des chats. Et d'autre part d'un algorithme d'apprentissage (comme ceux de "deep learning") en lui précisant les photos qui contiennent un chat et celles qui n'en contiennent pas.

Si on se concentre sur la construction du modèle (l'apprentissage automatique en question), l'algorithme va prendre chaque photo (avec ou sans chat) et rechercher une configuration du modèle qui évaluera positivement les photos avec un chat et inversement.

Pour mieux comprendre, faisons l'analogie avec certains jeux en bois : la configuration du modèle revient à fixer la position des plots en bois pour reconnaître les formes comme dans la photo ci-dessous.



L'apprentissage automatique s'apparente à la configuration des plots en bois pour identifier les formes.

Un modèle ne comprend pas ce qu'il évalue

L'apprentissage automatique est donc un algorithme qui produit un modèle sur la base de données fournies et évaluées par l'homme. Mais ce modèle (et par extension l'algorithme capable de l'utiliser) ne comprend pas (au sens de la construction et manipulation de concepts) le contenu des données soumises. Il est juste capable de corrélérer des signaux (issues des images) pour produire un score.

La magie des algorithmes par apprentissage automatique réside dans le choix des signaux à prendre en compte. C'est l'algorithme qui manipule très rapidement tous les signaux de toutes les photos pour trouver une bonne configuration. Par exemple, pensez à l'exercice qui consiste à trier des caisses de jouets. Certains modèles de classement se baseront sur la couleur, d'autres sur la taille ou encore sur la forme, etc. Le résultat est souvent surprenant car on ne peut pas raisonnablement expliquer le pourquoi et le comment de la configuration produite ce qui explique le côté "boîte noire" souvent associé à cette catégorie d'algorithmes.

La connaissance du concept de chat (et les relations avec les autres concepts comme ce qui le différencie d'un chien, etc.) n'est pas connu du programme.

Le programme n'évolue pas

L'algorithme qui permet de configurer un modèle est toujours le même (sauf si un humain vient à le faire évoluer). De même, le programme qui utilise le modèle n'évolue pas au fil de son utilisation.

Parfois on peut lire que "l'algorithme s'améliore avec le temps" mais il n'en est rien. Contrairement à la croyance populaire les algorithmes n'évoluent pas tous seuls. Si l'algorithme n'évolue pas, le modèle quant à lui peut être périodiquement mis à jour avec de nouvelles données. Cette mise à jour peut-être plus ou moins fréquente.

Par exemple, lorsque vous renseignez des CAPCHA sur Internet, vous donnez (à [Google](#) en l'occurrence) de nouvelles données pour améliorer son algorithme de reconnaissance de numéro des habitations.

L'homme reste toujours le gardien du sens

Les machines ne peuvent donc pas se perfectionner sans le concours des hommes. Les techniques d'apprentissage automatique permettent uniquement de reconnaître des patterns dans les données soumises. En aucun cas un programme utilisant ses techniques constitue une avancée sur la façon dont les machines peuvent comprendre le monde.

Si on devait se prêter à l'exercice périlleux de la prospective, nous pensons que l'avenir de l'intelligence artificielle sera du côté d'approches capables de concilier des modèles de connaissances (comme des ontologies et des terminologies) et les traitements statistiques et connexionnistes comme les réseaux de neurones. Plus particulièrement, les travaux les plus intéressants pour l'avenir de l'intelligence artificielle sont probablement ceux qui portent sur les terminologies c'est-à-dire des représentations formelles des concepts d'un domaine (une ontologie) et les termes pour en parler. On peut alors rêver d'un langage suffisamment formel (pour les machines) et riche sémantiquement (pour les humains) qui jouera l'interface idéal pour apporter du sens aux machines.

Mais d'ici là il faut se rendre à l'évidence, lorsque nous utilisons des techniques d'apprentissage automatique, nos machines calculent mais n'apprennent pas.

Christophe Tricot

A la suite de son doctorat en informatique appliquée à la modélisation et l'analyse sémantique de données, Christophe Tricot a développé son expertise en intelligence artificielle ces onze dernières années.

Après avoir occupé des postes d'ingénieur de recherche puis de chef de produit marketing, il a cofondé à partir de 2011 plusieurs sociétés de conseils et de développement de logiciels, telles que MetaCarto, Calopea et Kybelex.

Il a également créé une startup spécialisée dans la conception et l'implémentation de "bots" pour l'analyse de données complexes.

A 37 ans, Christophe Tricot rejoint aujourd'hui l'équipe Kynapse byOpen en tant que Manager pour développer une nouvelle offre "Intelligence artificielle".