

bas. Elle part du principe que, si l'on peut obtenir une connaissance abstraite à partir de données concrètes, c'est parce que l'on sait déjà beaucoup de choses et, en particulier, parce que le cerveau comprend déjà des concepts abstraits fondamentaux. Comme des scientifiques, nous pouvons utiliser ces concepts pour formuler des hypothèses sur le monde et faire des prédictions sur l'allure que devraient avoir les données (les événements) si ces hypothèses sont vérifiées – le contraire d'essayer d'extraire des motifs à partir des données brutes elles-mêmes, comme on le fait dans l'approche ascendante.

APPROCHE DESCENDANTE

On peut illustrer parfaitement cette idée en revenant sur la plaie des spams avec un cas réel auquel j'ai été confrontée. J'ai reçu un courriel de l'éditeur d'une revue au nom étrange, faisant référence spécifiquement à l'un de mes articles et me proposant d'écrire un article pour le périodique en question. Pas de Nigeria, pas de Viagra, pas de million de dollars, le courriel ne présentait aucun des indices trahissant généralement un spam. Mais en utilisant ce que je savais déjà et en réfléchissant de manière abstraite au processus qui produit les messages indésirables, j'étais en mesure de juger suspect ce message.

Pour commencer, je sais que les spammeurs tentent d'extorquer de l'argent aux gens en faisant appel à la cupidité humaine, et les universitaires peuvent être aussi avides de publier que

exemple, et j'ai pu ensuite tester mon hypothèse en vérifiant l'authenticité de l'éditeur par une requête sur un moteur de recherche.

Un informaticien qualifierait mon raisonnement de «modèle génératif», c'est-à-dire un modèle capable de représenter des concepts abstraits. Ce même modèle peut également décrire le processus utilisé pour produire une hypothèse, le raisonnement qui a conduit à la conclusion que le message pouvait être un simple spam. Ce modèle me permet d'expliquer comment fonctionne cette forme de courriel indésirable, mais il me permet également d'imaginer d'autres formes de spams, ou même un type différent de tous ceux que j'ai déjà vus ou dont j'ai entendu parler. Quand je reçois le message de la revue, le modèle me permet de travailler à rebours, de retracer pas à pas pourquoi cela doit être du spam.

Les modèles génératifs occupaient une place primordiale dans la première vague de l'intelligence artificielle et des sciences cognitives, dans les années 1950 et 1960. Mais ils avaient aussi leurs limites. Tout d'abord, la plupart des motifs discriminants pourraient, en principe, être expliqués par de nombreuses hypothèses différentes. Dans mon cas, cela pourrait être que le courriel était vraiment légitime, même si cela semble improbable. Ainsi, les modèles génératifs doivent incorporer des notions probabilistes, ce qui est l'un des développements récents les plus importants pour ces méthodes. En second lieu, on ne sait pas bien d'où viennent les concepts fondamentaux qui charpentent les modèles génératifs. Des penseurs tels que Descartes et Noam Chomsky ont suggéré que nous naissons avec, mais vient-on vraiment au monde en sachant que la cupidité mène aux escroqueries?

Les modèles bayésiens (l'un des principaux exemples récents de méthode descendante) tentent de traiter les deux questions. Portant le nom du statisticien et philosophe anglais du XVIII^e siècle Thomas Bayes, ils intègrent la théorie des probabilités aux modèles génératifs à l'aide d'une technique appelée inférence bayésienne. Un modèle génératif probabiliste peut vous indiquer la probabilité de voir un motif spécifique dans les données si une hypothèse particulière est vraie. Si le courriel est un message indésirable, il s'adresse probablement à la cupidité du lecteur. Mais bien sûr, un message pourrait aussi éveiller la cupidité sans être indésirable. Un modèle bayésien combine la connaissance que vous avez déjà sur les hypothèses vraisemblables avec les données dont vous disposez pour vous permettre de calculer, avec une assez bonne précision, la probabilité que le courriel soit légitime ou indésirable.

Cette méthode descendante correspond mieux que son homologue ascendante à ce que nous savons des mécanismes d'apprentissage ➤

Les modèles bayésiens sont des modèles d'apprentissage qui calculent des probabilités

L'individu lambda sera avide de gagner un million de dollars ou d'améliorer ses performances sexuelles. Je sais aussi que des revues légitimes «en accès libre» ont commencé à couvrir leurs coûts en faisant payer les auteurs plutôt que les abonnés. De plus, mes travaux n'ont rien à voir avec le titre de la revue. En regroupant toutes ces informations, j'ai émis l'hypothèse plausible selon laquelle il s'agissait d'un spam visant à convaincre des universitaires de payer pour «publier» un article dans une fausse revue. J'ai pu tirer cette conclusion à partir d'un seul

> chez l'enfant. C'est pourquoi, depuis quinze ans, mes collègues et moi utilisons les modèles bayésiens dans nos travaux sur le développement des enfants. Notre laboratoire et d'autres ont utilisé ces techniques pour comprendre comment les enfants apprennent les relations de cause à effet, et prédire comment et quand les jeunes changent de point de vue sur une conviction erronée.

Les méthodes bayésiennes sont également une excellente manière d'apprendre aux machines à apprendre comme les humains. En 2015, Joshua Tenenbaum, du MIT (l'institut de technologie du Massachusetts), Brenden Lake, de l'université de New York, et leurs collègues ont publié une étude dans laquelle ils ont mis au point un système d'intelligence artificielle capable de reconnaître des caractères manuscrits inhabituels, une tâche simple pour les humains mais extrêmement ardue pour les ordinateurs.

Pensez un instant à vos propres capacités de reconnaissance. Même si vous n'avez jamais vu un caractère sur un rouleau japonais, vous saurez sans doute dire s'il s'agit du même ou d'un autre que sur le rouleau voisin. Vous saurez probablement le dessiner, et même concevoir un faux caractère japonais, et également comprendre que son allure est très différente de celle d'un caractère coréen ou russe. C'est exactement ce que le logiciel de l'équipe de Joshua Tenenbaum est parvenue à faire.

Dans une méthode ascendante, on présenterait à l'ordinateur des milliers d'exemples, et il utiliserait les motifs repérés dans ces exemples pour identifier les nouveaux caractères. Au lieu de cela, le programme bayésien a donné à la machine un modèle général de tracé de caractère: par exemple, une barre peut aller vers la droite ou vers la gauche. Et une fois que le logiciel a terminé un caractère, il passe au suivant.

Quand le programme voyait un caractère donné, il pouvait inférer la séquence de traits nécessaire pour le tracer, et pouvait ensuite produire lui-même un jeu de traits similaire. Il le faisait de la même manière que j'ai inféré la série d'étapes qui avait conduit à mon message douteux envoyé par la pseudorevue. Au lieu d'évaluer la probabilité que ce message soit une escroquerie, le modèle de Tenenbaum devinait si une séquence particulière de traits avait des chances de donner le caractère souhaité. Ce programme descendant fonctionnait bien mieux que l'apprentissage profond appliqué aux mêmes données exactement, et il reproduisait assez fidèlement les performances des humains.

Ces deux approches majeures de l'apprentissage machine, ascendante et descendante, ont des forces et des faiblesses complémentaires. Avec la méthode ascendante, l'ordinateur n'a pas besoin d'avoir au départ la moindre compréhension de ce qu'est un chat, mais il a besoin de beaucoup de données. Le système bayésien, quant à lui, est capable d'apprendre à

partir de quelques exemples seulement, et il peut faire des généralisations plus larges. En revanche, cette approche descendante nécessite beaucoup de travail en amont pour articuler le bon jeu d'hypothèses. Et les concepteurs des deux types de systèmes peuvent rencontrer des obstacles similaires. L'une et l'autre approches ne fonctionnent que sur des problèmes relativement étroits et bien définis, comme la reconnaissance de caractères écrits ou de chats, ou les jeux sur console Atari.

LES ENFANTS EXPLOITENT LE MEILLEUR DES DEUX APPROCHES

Les enfants ne sont pas soumis aux mêmes contraintes. Les psychologues du développement ont trouvé que les jeunes combinent les meilleures qualités de chaque approche, ce qui les emmène beaucoup plus loin. Augie peut apprendre juste à partir d'un ou deux exemples, à la manière d'un système descendant. Mais il extrait aussi des corrélations à partir des données elles-mêmes, comme un système ascendant, et en déduit ainsi certaines idées qui n'étaient pas présentes initialement.

Augie a bien d'autres cordes à son arc. Il reconnaît immédiatement les chats et distingue les lettres, mais il peut également faire de nouvelles inférences créatives et étonnantes qui vont bien au-delà de son expérience ou de ses connaissances préalables. Il a récemment expliqué que si un adulte veut redevenir un enfant, il devrait essayer de ne pas manger de légumes sains, puisque ce sont eux qui font qu'un enfant grandit et devient adulte. Nous n'avons pratiquement aucune idée de la manière dont ce type de raisonnement créatif émerge.

Nous devrions nous rappeler les pouvoirs encore mystérieux de l'esprit humain quand nous entendons certains affirmer que l'intelligence artificielle est une menace existentielle. Intelligence artificielle et apprentissage automatique sont des termes qui font peur. Et à certains égards, cela est justifié – la recherche militaire travaille à utiliser ces systèmes pour commander des armes, par exemple.

Mais la stupidité naturelle peut faire plus de dégâts que l'intelligence artificielle, et nous autres humains devons faire preuve de beaucoup plus d'intelligence que nous avons pu le faire par le passé afin d'encadrer correctement ces nouvelles technologies. Même si les progrès de l'informatique résultent de l'augmentation de la puissance de calcul et des masses de données disponibles, plutôt que de révolutions conceptuelles dans notre compréhension de l'esprit, ils n'en peuvent pas moins avoir des conséquences pratiques capitales. Cela dit, nous n'avons pas de raisons de penser qu'un nouveau golem technologique est sur le point d'être lâché sur le monde. ■

BIBLIOGRAPHIE

Y. Bengio, **La révolution de l'apprentissage profond**, *Pour la Science* n° 465, juillet 2016.

A. Gopnik, **The Gardener and the Carpenter - What the new science of child development tells us about the relationship between parents and children**, Farrar, Straus and Giroux, 2016.

M. L. Brenden et al., **Human-level concept learning through probabilistic program induction**, *Science*, vol. 350, pp. 1332-1338, 2015.

A. Gopnik et al., **Bayesian networks, Bayesian learning and cognitive development**, *Developmental Science*, vol. 10(3), pp. 281-287, 2007.